

Université de Montréal

Expérience comportementale et modélisation par réseau neuronal des différences entre
les processus de catégorisation par règles logiques et par ressemblance familiale

Par

Laurence Morissette

Département de psychologie

Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures

en vue de l'obtention du grade de

Maître ès sciences (M. Sc.) en psychologie

Mai 2011

© Laurence Morissette

Université de Montréal
Faculté des études supérieures

Ce mémoire intitulé:

Expérience comportementale et modélisation par réseau neuronal des différences
entre les processus de catégorisation par règles logiques et par ressemblance familiale

présenté par :

Laurence Morissette

a été évalué par un jury composé des personnes suivantes :

Marc Schoenwiesner
président-rapporteur

Denis Cousineau
directeur de recherche

Sylvain Chartier
membre du jury

Mots clés: catégorisation implicite, catégorisation explicite, modélisation, réseaux neuronaux, transfert de l'apprentissage

Notre étude est bipartite. En premier lieu nous avons effectué une étude empirique des différences entre les processus de catégorisation explicite (verbalisable) et implicite (non-verbalisable). Nous avons examiné la difficulté et le temps nécessaire pour apprendre trois tâches de catégorisation dites par air de famille, par règle logique conjonctive et par règle logique disjonctive. Nous avons ensuite utilisé un réseau neuronal pour modéliser la catégorisation en lui faisant compléter les mêmes tâches. La comparaison entre les deux nous permet de juger de l'adéquation du modèle. Les données empiriques ont montré un effet de la typicité et de la familiarité en accord avec la documentation et nous trouvons que la tâche de catégorisation par règle disjonctive est la plus difficile alors que la tâche de catégorisation par air de famille est la plus facile. La modélisation par le réseau est une réussite partielle mais nous présentons des solutions afin qu'un réseau futur puisse modéliser le processus catégoriel humain efficacement

Keywords: implicit categorization, explicit categorization, modelization, neural networks, learning transfer

Our present research was twofold. First, we conducted a study of the differences in the categorization processes between explicit (verbalizable) and implicit (nonverbalizable) tasks. The differentiation was done in term of difficulty and time necessary to learn the rule of the category for family resemblance, conjunctive and disjunctive rules. We then used a neural network to model the categorization and had it complete the same tasks. The comparison between the two gave us insight into how (and if) the network can be used as a model of human categorization. The empirical data confirmed an effect of familiarity and typicality, as supported by previous studies, and we confirmed that the disjunctive task was the hardest to learn for humans while our results point toward the family resemblance task as being the easiest. The modelization by the neural network was partially successful at best but we present options that could permit a next generation neural network to model the categorization process truthfully.

Table des matières

Table des matières	v
Liste des figures	vii
Liste des tableaux	viii
Théorie classique	1
Théorie prototypiste	1
Théorie exemplariste	3
Théorie de la catégorisation par frontière de décision	4
Théories mixtes	5
Nature et complexité des catégories	6
Position du présent projet	7
Choix des tâches	7
Méthodologie	12
Section empirique	12
Sujets	12
Matériel	12
Procédure	15
Section modélisation	16
Matériel	16
Réseau	16
Architecture	17
Fonction de transmission	18
Règle d'apprentissage	18
Déroulement de la simulation	19
Résultats empiriques:	19
Phase d'apprentissage	19

Comparaison entre les deux phases.....	23
Phase de transfert	24
Résultats de la modélisation	27
Interprétation	34
Résultats empiriques	34
Problèmes méthodologiques	36
Résultats de la modélisation	37
Modifications à apporter au réseau.....	39
Modélisation future.....	41
Conclusion.....	43
Références	45
Annexe 1	I
Annexe 2	VII
Annexe 3	IX

Liste des figures

Figure 1 : Stimuli représentant la catégorie A pour les trois tâches	13
Figure 2 : Architecture du réseau	17
Figure 3 : Fonction de voisinage	18
Figure 4 : Taux de bonnes réponses durant la phase d'apprentissage et la phase de transfert	25
Figure 5 : Carte topologique des activations du réseau durant la phase d'apprentissage pour la tâche de catégorisation par air de famille.....	29
Figure 6 : Carte topologique des activations du réseau pour les stimuli avec une valeur nouvelle durant la phase de transfert pour la tâche de catégorisation par air de famille	31
Figure 7 : Carte topologique des activations du réseau pour les stimuli durant la phase de transfert pour la tâche de catégorisation par air de famille	31
Figure 8 : Carte topologique des activations du réseau durant la phase d'apprentissage pour la tâche de catégorisation par règle logique conjonctive	32
Figure 9 : Carte topologique des activations du réseau pour les stimuli durant la phase de transfert pour la tâche de catégorisation par règle logique conjonctive.....	32
Figure 10 : Carte topologique des activations du réseau durant la phase d'apprentissage pour la tâche de catégorisation par règle logique disjonctive.....	33
Figure 11 : Carte topologique des activations du réseau pour les stimuli avec une valeur nouvelle durant la phase de transfert pour la tâche de catégorisation par règle logique disjonctive.....	34
Figure 12 : Carte topologique des activations du réseau pour les stimuli durant la phase de transfert pour la tâche de catégorisation par règle logique disjonctive	34
Figure 13 : Test sur le pattern d'activations 4×4 avec 4 activations par couche	VII
Figure 14 : Test sur le pattern d'activations 4×4 avec 8 activations par couche	VII
Figure 15 : Activations pour la tâche de catégorisation par règle conjonctive avec trois attributs.....	VIII

Liste des tableaux

Tableau IV : Nombres de blocs nécessaires à l'apprentissage des tâches.....	20
Tableau VI : Taux de bonnes réponses et temps de réponse par groupes de familiarité dans les tâches de catégorisation par règle conjonctive et disjonctive	22
Tableau VII : Changements des temps de réponse et des taux de bonnes réponses entre le dernier bloc de la phase d'apprentissage et la phase de transfert	23
Tableau VIII : Temps de réponse et taux de bonnes réponses durant la phase de transfert	24
Tableau IX : Taux de bonnes réponses pour les différents niveaux de familiarité durant la phase de transfert de la tâche de catégorisations par règle disjonctive	27
Tableau X : Paramètres d'apprentissage des attributs uniques.....	28
Tableau XI : Paramètres des trois tâches de catégorisation effectuées par le réseau	28
Tableau XII : Résultats de l'apprentissage par le réseau de la tâche de catégorisation par air de famille.....	30
Tableau I : Valeur des attributs des stimuli présentés durant la tâche de catégorisation par air de famille.....	I
Tableau II : Valeur des attributs des stimuli présentés durant la tâche de catégorisation par règle conjonctive	III
Tableau III : Valeur des attributs des stimuli présentés durant la tâche de catégorisation par règle disjonctive	IV
Tableau V : Moyennes durant le dernier bloc de la phase d'apprentissage	VI

La catégorisation est un mécanisme polyvalent qui permet aux humains de généraliser, d'exclure l'information non-pertinente, d'unifier les attributs, et à un niveau plus fondamental, de conceptualiser et d'acquérir des connaissances (Anderson, 1991, Medin, 1989). Les catégories peuvent avoir une influence sur le langage, sur la résolution de problèmes, sur les processus mnésiques; elles occupent donc une place unique dans la cognition humaine. C'est pourquoi comprendre le ou les processus responsables de l'acquisition et de l'utilisation de ces concepts ainsi que leur nature même a inspiré les travaux de chercheurs en psychologie cognitive depuis plusieurs décennies.

Théorie classique

La première théorie formelle de la catégorisation a été développée par Smith et Medin (1981). La théorie classique de la catégorisation, comme elle est maintenant connue, pose certains postulats concernant les catégories. Le principal stipule que les concepts sont représentés mentalement comme des définitions qui donnent les attributs nécessaires et suffisants pour décrire l'appartenance à une catégorie. Il n'existe donc pas de possibilité d'ambiguïté quant à l'inclusion d'un item dans une catégorie, ni concernant le degré de représentativité du dit item par rapport à la catégorie; tous les items appartenant à la catégorie sont d'égale valeur pour la représenter. Puisque ces postulats ne concordent pas avec les observations empiriques faites, plusieurs autres théories ont été développées pour tenter d'expliquer ces observations. Les trois plus grandes sont présentées plus avant: soit respectivement le prototypisme (apprentissage par formation d'un prototype unique), l'exemplarisme (apprentissage par accumulation d'exemplaires) et la théorie de catégorisation par frontière de décision.

Théorie prototypiste

La théorie prototypiste (Rosch 1975, Minda & Smith 2001, 2002) part de la prémisse que l'on crée une catégorie en maintenant en mémoire un ensemble pondéré des caractéristiques moyennes représentant la catégorie. Il n'y a donc qu'une seule liste d'attributs qui représente la catégorie. Cela implique que pour qu'un jugement puisse être posé sur les caractéristiques d'un attribut, chaque attribut doit pouvoir être évalué indépendamment des autres.

La catégorisation d'un nouvel item se fait en comparant les attributs de l'item aux listes d'attributs des catégories auxquelles il est possible pour cet item d'appartenir. L'item gagne des points pour chaque attribut présent et en perd soit pour chaque attribut appartenant au prototype qui est absent de l'item ou pour chaque attribut présent pour l'item mais absent de la liste de la catégorie. L'item est alors catégorisé comme appartenant à la catégorie pour lequel il possède le plus d'attributs pertinents. C'est pourquoi même si certains attributs sont absents mais que d'autres attributs sont cohérents avec l'appartenance à la catégorie, l'item peut quand même être catégorisé correctement. Ce dernier élément est très important quant à la plausibilité réelle de ce modèle, car il est rare que les items vus dans la réalité répondent à tous les critères de la catégorie. Pour qu'un item soit considéré typique de la catégorie, il doit être semblable au prototype. En conséquence, plus l'item possède d'attributs appartenant à la liste définissant la catégorie, et ce calculé de façon additive, et plus il est considéré typique de la catégorie (Laroche, Cousineau & Archambault 2006).

Rosch (1975) a défini le type de catégorisation présentée à l'instant comme une catégorisation par air de famille. Selon la chercheuse, les catégories naturelles ('naturel' est opposé ici à 'artificiel') seraient formées de cette façon. Barsalou (1985, 1987) a raffiné cette affirmation par des études montrant que les catégories seraient plutôt une hiérarchie où les items les plus typiques sont plus facilement récupérables et les items de moins en moins typiques sont de moins en moins facilement récupérable. D'autres chercheurs (Cohen & Murphy 1984) ont proposés que ces attributs seraient non pas présents sous la forme d'une simple liste mais bien arrangés selon un schéma pour lequel les relations entre les différents attributs sont définies.

Un des écueils rencontré par le modèle prototypiste est qu'il est très difficile de regrouper dans un seul élément conceptuel des pondérations d'attributs représentant tous les items appartenant à la catégorie, surtout si cette catégorie comprend des items très disparates et ayant des attributs contradictoires. Il en découle que la théorie prototypiste ne peut pas expliquer l'apprentissage d'une catégorie qui est régie par une règle non-verbalisable disjonctive, puisqu'il est impossible de créer un prototype comprenant avec

poids égal des caractéristiques opposées. Il peut par contre être argumenté que les règles disjonctives sont arbitraires et ne découlent pas directement des attributs des items vus.

Théorie exemplariste

Pour la théorie exemplariste (Medin & Schaffer 1978, Nosofsky 1988), la catégorisation d'un nouvel item se fait par la comparaison entre cet item observé et les exemplaires retenus en mémoire. Ces exemplaires sont composés de tous les items ayant été observés et ensuite classés comme appartenant à une catégorie, alors que la comparaison est un calcul de la ressemblance entre l'item nouveau et tous les items connus. La catégorie contenant l'item le plus semblable à l'item nouveau est choisie comme la catégorie d'appartenance de l'item. Ici, la typicité d'un item à sa catégorie est proportionnelle à sa similarité à tous les autres items mémorisés de la catégorie considérée.

Il n'y a par contre pas consensus à savoir si chaque apparition d'un item représente un exemplaire (Nosofsky 1988) ou si la présentation d'un item ne compte qu'une fois. Barsalou, Huttenlocher et Lamberts (1998) affirment d'ailleurs que puisqu'il est parfois difficile de faire la différence entre le même item présenté plusieurs fois et des items différents mais semblables, c'est la fréquence de l'item tel qu'interprété comme étant le même qui augmente le jugement de typicité de cet item. Pour ces chercheurs, le calcul de similarité serait fait de façon multiplicative et non pas additive, de quoi il découle qu'il est préférable d'avoir une grande similarité avec quelques items plutôt qu'une similarité moyenne avec plusieurs items.

Murphy (2004) présente l'hypothèse que les effets de catégorisation exemplariste seraient équivalents aux effets de mémoire implicite, en ce sens que le choix effectué dans la catégorisation, si ce choix est fait par ressemblance avec un autre item déjà vu, est rarement conscient. Il est en effet très improbable de pouvoir rappeler de façon explicite quel item a dirigé notre choix de catégorie.

Pour la théorie exemplariste, l'apprentissage de catégories pouvant être apprises par règles ne sont pas différentes des tâches d'apprentissage de catégories par

ressemblance familiale. Il n'y a donc pas de difficulté supplémentaire d'un type de catégorisation (implicite versus explicite) pour la théorie.

Théorie de la catégorisation par frontière de décision

Les deux théories présentées précédemment ont connus du succès dans la prédiction et/ou l'ajustement à des données empiriques. Leur puissance d'ajustement peut s'expliquer par la complexité des modèles formels utilisés (grand nombre de paramètres, métriques hautement non-linéaires, etc.). Néanmoins, plusieurs de leurs prédictions ont été vérifiées expérimentalement. Face à une telle situation où il ne semble pas possible de départager une meilleure approche, Ashby et ses collègues ont choisi d'ignorer la dichotomie exemplaire-prototype pour aborder le problème sous un autre angle.

Leur théorie, supposant elle aussi une catégorisation par un seul système, est la théorie de la catégorisation par frontière de décision (*decision bound*, Maddox & Ashby 1993, 1994), qui découle de la théorie générale de la reconnaissance d'Ashby et Townsend (1986). Au lieu de comparer un ou des items ensemble, cette théorie affirme que l'espace des attributs est séparé par des frontières et que tous les items à l'intérieur d'une frontière donnée appartiennent à la catégorie. L'apprentissage de la catégorie est alors l'apprentissage de la frontière ou des régions associées à chaque côté des frontières. La catégorisation d'un nouvel item revient à trouver entre quelles frontières l'item est situé. Les auteurs rajoutent un postulat à la théorie qui la dissocie des deux précédemment présentée : la présence d'une variabilité de l'information perceptuelle associée à chaque stimulus pour chaque instance du stimulus vue. Cette variabilité permet à la théorie d'affirmer une adéquation de processus entre la catégorisation et l'identification dans le contexte de la théorie, les deux processus nécessitant l'apprentissage de percepts reliés à une certaine réponse. Dans la théorie, cinq éléments limite la précision de la réponse: le bruit perceptuel, la sélection d'une frontière catégorielle sous-optimale, la variabilité dans la récupération en mémoire de la frontière, le biais de réponse et la variabilité dans le rappel du critère de réponse (Ashby & Maddox, 1993). La typicité d'un item peut alors être vue comme la distance d'une frontière : plus l'élément s'en éloigne et moins il y a d'ambiguïté sur son appartenance à cette catégorie. La théorie du *decision bound* fait une différence entre les frontières linéairement séparables, qui sont équivalentes aux règles

verbalisables, et celles non-linéairement séparables, qui ne peuvent pas être formulée par une règle simple.

Théories mixtes

Ces trois théories sont toutes unitaires en ce sens qu'elles expliquent tous les processus de catégorisation par un seul processus. Cette position est bien entendue extrême et plusieurs approches mixtes ont été proposées, mélangeant plus d'un processus de catégorisation en parallèle (e.g. RULEX, Nosofsky, Palmeri & McKinley 1994), utilisant plus d'un niveau dans l'architecture (e.g. EBRW, Nosofsky & Palmeri, 1997), implémenté par une architecture neuronale (e.g. ATRIUM, Erickson & Kruschke, 1998 ou SUSTAIN, Love, Medin & Gureckis, 2004) ou par un modèle d'échantillonnage (Lamberts, 2000).

Ashby et Maddox (2005) ont plus récemment posés l'hypothèse de systèmes distincts pour les catégories pouvant être décrites explicitement par des règles verbalisables et celles ne pouvant pas l'être. Leur modèle, COVIS (*COmpetition between Verbal and Implicit Systems*, Ashby, Alfonso-Reese, Turken, & Waldron, 1998) propose la présence de deux systèmes distincts. Le premier, le système explicite, se charge de la catégorisation selon des règles pouvant être verbalisées et procède selon un raisonnement conscient. Il nécessite une séparabilité des attributs composant les stimuli et une possibilité de porter attention de façon sélective sur chacun. Sa tâche principale est de choisir la règle pertinente parmi les types de règles verbales et d'accéder à l'information visuelle et sémantique appropriée. Ce système fait aussi usage de la mémoire de travail, de l'attention et des fonctions exécutives. Le deuxième, le système implicite, se charge de la catégorisation selon des règles non verbalisables (aussi appelée tâche d'intégration de l'information), et utilise l'apprentissage procédural. Il est une implémentation directe de la théorie de la catégorisation par *décision bound* d'Ashby et al. Sa tâche principale est de trouver la frontière de décision optimale permettant de séparer l'espace des stimuli. Le COVIS suppose donc une différence intrinsèque selon la nature même du stimulus à évaluer quand celui-ci peut être catégorisé de façon optimale par l'application d'une règle verbalisable et lorsque ceci n'est pas possible.

Nature et complexité des catégories

Une approche pour mieux restreindre le champ d'étude est d'avoir une meilleure compréhension de la nature des catégories proprement dites. Puisque les catégories n'ont pas de substrat physique, il s'agit donc d'étudier les caractéristiques logiques des catégories possibles.

Neisser et Weene en 1962 ont été parmi les premiers à étudier la complexité des règles définissant les catégories. Ils ont trouvé 3 niveaux de règle, chacun contenant des règles logiques d'égale difficulté, et chaque niveau étant plus complexe que le précédent. Au premier niveau se retrouvent la présence (a) et l'absence d'attributs (non-a). Au deuxième niveau se trouvent la conjonction (a et b), la disjonction (a ou b ou les deux), l'exclusion (a et non b), l'absence disjonctive (ni a ou ni b ou ni a et ni b), l'absence conjonctive (ni a ni b) et l'implication (si a, b). Au troisième et dernier niveau se trouvent le ou exclusif (a ou b mais pas a et b) et le l'un-et-l'autre/l'un-ni-l'autre (a et b ou non a et non b).

Feldman, un autre pionnier dans ce domaine, a examiné la complexité de toutes les règles logiques possibles avec quatre attributs (Feldman, 2000, 2003, révisé par Lafond, Lacouture & Mineau, 2007). Ses travaux permettent de classifier la difficulté subjective des catégories lorsqu'elles sont vues comme le résultat de l'application d'une règle logique. Feldman affirme que cette difficulté subjective est directement proportionnelle à sa complexité booléenne, soit la longueur de la plus courte proposition logique équivalente. Pour pouvoir parler d'un attribut booléen, celui-ci doit pouvoir ne prendre que deux valeurs. Lafond et al. ont confirmé expérimentalement que la complexité subjective des règles conjonctives est toujours moindre que la complexité des règles disjonctives utilisant le même nombre d'attributs.

Position du présent projet

Le présent projet est divisé en deux parties. Nous n'avons pas trouvé dans la documentation d'expérience qui visait à comparer la difficulté subjective des tâches de catégorisation par ressemblance familiale et par règle logique. C'est pourquoi dans la première phase du projet nous avons effectué une comparaison des processus de catégorisation explicites et implicites de façon expérimentale afin de distinguer les différences qualitatives et quantitatives entre les deux. Dans la seconde phase, nous avons utilisé un réseau neuronal afin de tenter de reproduire les résultats expérimentaux et de modéliser les processus catégoriels humains observés empiriquement.

Choix des tâches

Les trois tâches que nous avons retenues ont été choisies de façon à couvrir les deux types de tâches de catégorisation trouvées les plus fréquemment dans la littérature. Premièrement, une tâche de catégorisation par ressemblance familiale où les items ont un niveau d'appartenance graduel, dans l'esprit des stimuli de Rosch. Nous voulions aussi incorporer des tâches par règle de décision sur le modèle de l'ensemble de stimuli de Shepard, Hovland et Jenkins (1961). Dans ce type de tâche, plusieurs types de règles sont souvent mis en comparaison. Nous avons d'abord choisi la règle conjonctive car il s'agit de la règle la plus simple après celle où il n'y a qu'un attribut unique. De façon complémentaire, nous avons choisi la règle disjonctive car cette règle est reconnue comme étant difficile à apprendre pour l'être humain comme pour les réseaux et nous permettrait donc de mettre en relief aisément la capacité de l'un et l'autre. Nous avons aussi choisi que les catégories ne se chevauchent pas pour la tâche de catégorisation par air de famille afin d'obtenir une correspondance avec les tâches de catégorisation par règle logique ou cette ambiguïté est impossible à implémenter. Cela nous permet aussi d'avoir une dissociation des catégories plus facile. Le fait que pour chaque tâche, les stimuli étaient séparé en 2 catégories correspond au paradigme utilisé le plus fréquemment en catégorisation et permet une interprétation facile des résultats obtenu avec le réseau quant à la position des attracteurs.

Caractéristiques recherchées dans le réseau

Afin de choisir un réseau ayant les caractéristiques correspondant le mieux à un modèle de la catégorisation humaine, plusieurs réseaux ont été envisagés.

Il existe deux stratégies dirigeant l'apprentissage des réseaux, la première étant de superviser l'apprentissage du réseau et la seconde de permettre au réseau d'apprendre par lui-même. Les réseaux supervisés ont été les premiers à être développés et les réseaux les plus simples appartiennent à cette catégorie. Le perceptron (Rosenblatt 1958), connu comme le premier réseau neuronal ainsi que le plus simple, a une règle d'apprentissage dirigée par la différence entre l'output obtenu par le réseau et l'output attendu. Les réseaux non supervisés à l'inverse n'ont pas d'output attendu vers lequel ils doivent tendre. Leur output est entièrement déterminé par l'information pouvant être extraite de l'input.

Les études neurologiques du cerveau ne permettent pas d'identifier une connaissance innée des structures catégorielles. De plus, il n'existe pas d'homuncule confirmant ou infirmant l'appartenance d'un item à telle ou telle catégorie dans le système visuel humain. L'apprentissage est plutôt fait par essai et erreur et par l'interaction dynamique entre le milieu et les réponses données. Le réseau permettant la meilleure correspondance avec le processus catégoriel humain est donc un réseau auto-apprenant.

Une deuxième distinction importante pour les réseaux est la direction de la transmission. Pour les réseaux les plus simples, l'information ne se propage que dans une seule direction (*feedforward*). La majorité des réseaux actuels plus élaborés permettent au résultat d'influencer les éléments plus avant dans le système. Ceci a une correspondance dans le processus humain dans le fait que les processus *bottom-up* et *top-down* influencent l'apprentissage.

Une troisième caractéristique espérée du réseau est que comme nous nous intéressons à la catégorisation et non la capacité d'identification de chacun des items, nous donnons priorité au processus de réduction de l'information en éléments invariants à

l'intérieur de la catégorie. Le réseau doit donc être en mesure de prendre l'information qui lui est transmise et en extraire les attributs discriminants inter-catégories.

La classe générale des réseaux à laquelle nous nous sommes intéressés est celle des réseaux neuronaux récurrents (RNN). Ceux-ci ont une mémoire interne leur permettant d'apprendre une séquence arbitraire d'inputs. Les mémoires récurrentes auto-associatives (RAM) sont une version non-supervisée de cette classe de réseaux. Les RAMs peuvent enregistrer des patrons qui sont corrélés entre eux et à travers une boucle de rétroaction, filtrer le bruit, compléter un patron incomplet et généraliser à des patrons nouveaux mais semblables. Cette dernière compétence est appelée plasticité et nous permet d'intégrer au réseau la plasticité catégorielle observée empiriquement chez les humains (Bornstein et al. 1976, Bornstein 1987, Pevtsov & Harnad 1997). Pour chaque patron, le réseau apprend un attracteur, qui peut être vu comme un prototype invariant. Par contre, comme ils enregistrent ces attracteurs comme des items sans bruit, les RAMs ne représentent pas un modèle psychologique du processus d'acquisition de catégorie humain, puisque plausiblement, le système visuel ajoute toujours un peu de bruit à ce qui est observé. De plus, le processus d'encodage des éléments perçus visuellement doit regrouper les items vus qui sont tous un peu différents mais appartenant essentiellement à une seule et même catégorie les regroupant tous. Finalement, les RAMs ne peuvent pas associer plusieurs inputs à une seule représentation (*many-to-one*), ce qui est très problématique pour une modélisation de la catégorisation.

Les mémoires bidirectionnelles associatives (BAM) de Kosko (1988) sont un type de réseau ayant une meilleure correspondance avec le fonctionnement humain. Elles sont dites hétéroassociatives, au sens qu'elles peuvent associer deux patrons distincts. Il devient donc possible d'associer un input avec une catégorie, quoique seulement de façon supervisée. Chartier et Boukadoum (2006) ont présenté un réseau sur le modèle du BAM, mais permettant d'associer des inputs à valeurs continues et pouvant faire des associations *many-to-one*.

Une dernière caractéristique recherchée dans le modèle est de pouvoir représenter de façon contiguë dans l'espace des exemplaires les items semblables. Kohonen a

proposé l'idée d'une carte topologique en 1984, dans son livre *Self-organization and associative memory*. Ce réseau a la capacité de faire du *clustering*; il apprend de façon non supervisée à créer une représentation (la carte) d'un stimulus multidimensionnel complexe à l'aide d'un nombre de dimensions moindres, tout en préservant les propriétés topologiques des inputs. La carte est composée d'unités simples reliées à l'input par des poids. Pour ce réseau, chaque unité de la carte représente regroupement, c'est-à-dire une unité qui sera toujours activée maximale par la présentation d'un certain type d'input (Chartier, Giguère & Langlois, 2009) de sorte que chaque item entrant est associé à l'un d'eux, dans un processus que l'on pourrait équivaloir à de la catégorisation. Le réseau de Kohonen a un certain nombre de lacunes : d'abord la métrique est celle du monde physique, ce qui est problématique pour un réseau neuronal évoluant sur des forces de connexions; ensuite la règle d'apprentissage est très sensible et difficile à ajuster pour garantir l'apprentissage (Cousineau, 2009) et le réapprentissage est impossible (Chartier, Giguère & Langlois, 2009), le réseau se figeant à la fin d'une première phase d'apprentissage.

Cependant, il existe un autre réseau, basé sur une règle d'apprentissage hébbien qui implémente une carte topologique. Ce réseau, *Feature Extraction with a Bidirectional Associative Memory*, FEBAM, est en mesure d'apprendre de façon non supervisée. Il a été développé par Sylvain Chartier en 2006 et a été raffiné depuis de façon à avoir une plus grande correspondance avec le fonctionnement présumé de la catégorisation et de l'identification humaines (Chartier, Renaud & Boukadoum, 2008, Chartier, Boukadoum, & Amiri, 2009, Chartier, Giguère & Langlois 2009). Ce réseau apprend des patrons d'inputs (des stimuli) en développant des attracteurs. Les unités voisines sont aussi sensibles à l'input, mais dans une moindre mesure, celles-ci étant mieux adaptées pour détecter des « morphes », des versions intermédiaires entre deux attracteurs, selon leur distance aux différents attracteurs. On peut donc dire que ce modèle est sensible aux exemplaires. Par l'existence de ces unités intermédiaires, le réseau est en mesure de généraliser et est résistant au bruit sur les items. On peut donc dire qu'il est à la fois stable et plastique. De plus, on peut obtenir une carte topologique de la position où les items appris sont situés les uns par rapport aux autres et ainsi savoir quels sont ceux étant considérés comme les plus semblables. Comme il utilise une métrique *k-winner-take-all* (kWTA), il implémente

aussi une forme de seuil, un input qui n'éveille pas suffisamment d'unités étant considéré ambiguë par le réseau. La règle kWTA sélectionne les k activations les plus fortes et ne considère que celles-là pour l'attribution des connexions entre inputs et outputs. Un élément peut être positionné entre 2 attracteurs et donc avoir une appartenance aux deux groupements, avec une force plus ou moins grande. Ce réseau entoure donc chaque attracteur d'une frontière catégorielle, le rendant semblable à la théorie de *decision bound*. Comme ce modèle est autoassociatif et que les paramètres d'apprentissages sont tous fixés uniquement en fonction de la taille de la couche d'entrée et de la taille de la carte topologique, le modélisateur a peu de liberté pour changer le comportement du réseau. Il y a donc peu de risque d'avoir un *overfitting* des données empiriques.

Le FEBAM-SOM, un réseau combinant les caractéristiques des BAM et des SOM a donc été choisi, afin de pouvoir utiliser à la fois la récurrence et le positionnement sur une carte topologique. Le cœur du modèle utilise une dynamique non-linéaire pour développer des attracteurs, chaque attracteur étant associé à une catégorie. Ces attracteurs évoluent dans une carte topologique à deux dimensions, ce qui permettra de voir les relations entre les catégories.

Hypothèses

Au cours du volet empirique proposé dans ce projet, nous nous attendons à trouver que la tâche d'apprentissage où les catégories sont définies à l'aide d'une règle conjonctive sera la mieux réussie et celle apprise la plus rapidement, suivie de la tâche d'apprentissage par ressemblance familiale et finalement la tâche d'apprentissage par règle disjonctive. Nous nous attendons à un transfert des apprentissages équivalents pour les trois tâches lors du passage à la phase de transfert, puisque les catégories devraient avoir été bien apprises à ce point. Nous prévoyons aussi répliquer les effets de la typicité et de la familiarité documentés dans la littérature.

Pour la modélisation à l'aide du réseau FEBAM-SOM, nous nous attendons à trouver une répartition des items dans l'espace topologique qui sont séparable linéairement pour les tâches de catégorisation par règle logique. Nous nous attendons à trouver un attracteur bien défini pour la catégorie A de la tâche conjonctive mais pas

d'attracteur pour la catégorie Non-A puisqu'il n'existe pas de structure propre à cette catégorie. Nous nous attendons à trouver quatre attracteurs pour la tâche disjonctive puisque le réseau non-supervisé n'a pas de façon d'apprendre à regrouper de façon arbitraire des stimuli ayant des caractéristiques opposées. Pour la tâche de catégorisation par ressemblance familiale, nous nous attendons à un continuum basé sur la ressemblance au prototype de l'une ou l'autre catégorie.

La relation ordinale entre les temps de réponses des différentes catégories chez les sujets humains devraient correspondre à la relation ordinale du nombre d'itérations nécessaire au réseau pour faire l'apprentissage.

Méthodologie

Section empirique

Sujets

48 étudiants de l'Université de Montréal ont complété l'expérience. Les résultats de 40 étudiants (26 femmes et 14 hommes) ont pu être analysés, les huit autres ayant soit réussi une seule tâche ou ayant un profil de réponse très différent de la moyenne. 19 provenaient de programmes de science naturelles ou de génie et 21 provenait de programmes de sciences humaines, de sciences sociales ou de lettres.

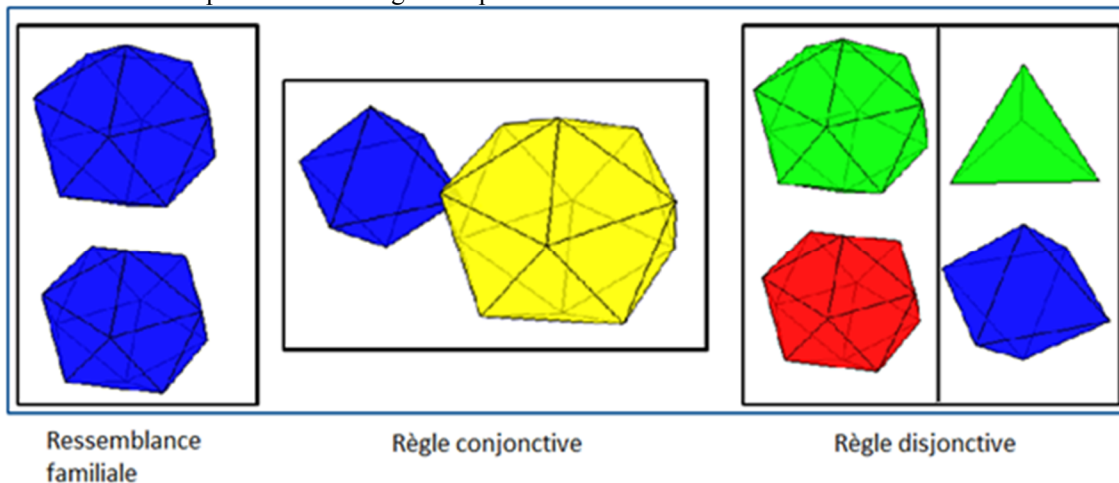
Matériel

L'expérience comporte trois ensembles de stimuli représentant deux polyèdres côte-à-côte, présentés sur un écran d'ordinateur. Quelques exemples de ces stimuli se trouvent à la Figure 1.

Les stimuli comporte cinq attributs, soit le nombre de côtés du 1^{er} polyèdre, le nombre de côtés du 2^e polyèdre, la couleur du 1^{er} polyèdre, la couleur du 2^e polyèdre et la position des deux polyèdres l'un par rapport à l'autre. Chacun des attributs peut prendre 4 valeurs différentes.

Les stimuli ont des attributs descriptibles verbalement, contrairement aux stimuli continus de type “dot pattern” (Homa 1987), car nous voulions utiliser les mêmes stimuli pour toutes les tâches et les stimuli continus sont peu appropriés pour l’utilisation dans des tâches de catégorisation par règle logique. En effet, l’intérêt même de ces stimuli est de mettre en relation de façon implicite les caractéristiques présentes dans le stimulus.

Figure 1
Stimuli représentant la catégorie A pour les trois tâches



L'expérience est composée des trois tâches préalablement présentées: une tâche de catégorisation par ressemblance familiale, une tâche de catégorisation par règle logique conjonctive et une tâche de catégorisation par règle logique disjonctive. Chacune des tâches est composée de deux phases: une phase d'apprentissage et une phase de transfert.

Dans la tâche de catégorisation par ressemblance familiale, les stimuli appartiennent à une de deux catégories. La catégorie d'appartenance est celle pour laquelle ils partagent le plus d'attributs communs avec le prototype de la catégorie. Au cours de la phase d'apprentissage, 24 stimuli sont présentés, 12 pour la catégorie A et 12 pour la catégorie B (présentés dans le Tableau I de l'Annexe 1). Pour chacun de ces 12 stimuli, six partagent quatre attributs avec le prototype et six en partagent trois. Cette manipulation vise à moduler la typicité des stimuli présentés. Durant cette phase, seulement trois des quatre valeurs possibles pour les attributs sont utilisées. Pendant la phase de transfert, 48 nouveaux stimuli sont répartis également entre les deux catégories

(aussi présentés dans le Tableau I). Des 24 stimuli de chaque catégorie, 12 incluent la 4^e valeur possible pour les attributs, de façon à intégrer un élément de nouveauté. Ces stimuli sont séparés en trois groupes de quatre stimuli ayant respectivement quatre, trois et deux attributs communs avec le prototype. Les 12 derniers stimuli sont des stimuli nouveaux mais n'intégrant pas la nouvelle valeur. Ils ont une similarité plus ou moins grande avec les stimuli présentés durant la phase d'apprentissage : chaque groupe de quatre stimuli partagent respectivement quatre, trois et deux valeurs d'attributs avec les items initiaux. Pour les deux premiers groupes, deux stimuli ont quatre valeurs communes avec le prototype de la catégorie, un en a trois et un en a deux, alors que pour le dernier groupe, un stimulus a 4 valeurs communes avec le prototype, deux en ont trois et un en a deux.

La règle régissant l'appartenance à la catégorie A durant pour la deuxième tâche de catégorisation est une règle conjonctive. Il s'agit d'une règle où 2 des 5 attributs sont pertinents à la catégorisation. Cette règle a un niveau de complexité de 2 selon la nomenclature de Feldman (2000), ce qui permet de la situer comme une règle de complexité basse. Cette tâche a une structure de type A/Non-A. Au cours de la phase d'apprentissage, 24 stimuli sont présentés, douze pour la catégorie A et douze pour la catégorie Non-A (présentés dans le Tableau II de l'Annexe 1). La règle logique est définie par la présence de deux valeurs spécifiques pour les deux premiers attributs (1 et 1 pour les A). Une manipulation visant à moduler la familiarité est effectuée sur les attributs 3 et 4. En effet, une valeur spécifique est présente neuf fois sur douze pour l'attribut trois (la valeur 2; condition de haute familiarité) et une autre valeur est présente six fois sur douze pour l'attribut quatre (la valeur 3; condition de moyenne familiarité) dans les stimuli appartenant à la catégorie A. Un item peut donc à la fois posséder une valeur hautement familière et une valeur moyennement familière puisque ce sont des attributs indépendants. Durant cette phase, les attributs quatre et cinq n'utilisent que trois des quatre valeurs possibles. Les attributs des stimuli n'appartenant pas à la catégorie A ont des valeurs équiprobables. La phase de transfert présente 16 nouveaux stimuli (présentés aussi dans le Tableau II) répartis également entre les deux catégories. Les stimuli font usage des valeurs n'ayant jamais été vues dans la phase d'apprentissage pour les attributs quatre et cinq pour intégrer un élément de nouveauté.

La troisième tâche est une tâche de catégorisation pour laquelle il existe deux catégories. La règle régissant l'appartenance aux catégories est une règle disjonctive où 2 attributs sont pertinents pour la catégorisation. Cette règle a un niveau de complexité de 4 selon la nomenclature de Feldman (2000), ce qui permet de la situer comme une règle de complexité élevée. Au cours de la phase d'apprentissage, 24 stimuli sont présentés, 12 pour la catégorie A et 12 pour la catégorie B (présentés dans le Tableau III de l'Annexe 1). La règle logique est définie par la présence de valeurs spécifiques pour les deux premiers attributs (1 et 0 ou 0 et 1 pour les membres de la catégorie A). Une manipulation visant à moduler la familiarité est effectuée sur les attributs 3 et 4. En effet, une valeur spécifique est présente neuf fois sur douze pour l'attribut trois (condition de haute familiarité) et une autre valeur est présente six fois sur douze pour l'attribut quatre (condition de moyenne familiarité) dans les stimuli appartenant à chaque catégorie. Durant cette phase, l'attribut quatre n'utilise que trois des quatre valeurs possibles. La phase de transfert présente 48 nouveaux stimuli (aussi présentés dans le Tableau III) répartis également entre les deux catégories. Des 24 items appartenant à chaque catégorie, 12 items présentent une valeur de l'attribut quatre non présentée durant la phase d'apprentissage, de façon à créer de la nouveauté. Dans les douze stimuli restants, six partagent trois valeurs et six partagent quatre valeurs avec les items déjà présentés.

Procédure

Les participants sont assis seuls dans une pièce ne contenant que l'ordinateur afin d'éviter les distractions. Les instructions sont présentées à l'écran et informent les participants de la séparation de l'expérience en 3 tâches, elles-mêmes séparées en 2 phases. Ils sont informés qu'ils recevront une rétroaction durant la phase d'apprentissage de chaque tâche et qu'ils doivent continuer cette phase jusqu'à l'obtention d'un taux de réponse de 90%, après quoi une phase de transfert sans rétroaction suit. Chaque rétroaction est d'une durée de 2 secondes et informe le participant sur la véracité de sa réponse et lui donne la catégorie appropriée s'il s'est trompé. Les sujets doivent observer les stimuli sur l'écran d'ordinateur et répondre grâce au clavier. Tous les participants doivent effectuer les 3 tâches qui composent l'expérience. Afin de contrôler les effets de pratique, 18 sujets complètent l'expérience dans l'ordre : tâche de catégorisation par air de

famille - tâche de catégorisation par règle conjonctive - tâche de catégorisation par règle disjonctive et 22 dans l'ordre inverse. Au cours de la phase d'apprentissage, pour chaque bloc d'essai, le sujet voit les 24 items composant les stimuli de la tâche dans un ordre aléatoire. Pour chaque stimulus présenté, le participant doit classer le stimulus comme appartenant à la catégorie A ou B pour les tâches 1 et 3 et A/Non-A pour la tâche 2 et répondre en appuyant sur une touche appropriée sur le clavier.

Section modélisation

Matériel

L'ensemble des stimuli présentés aux sujets humains sont d'abord codés en une forme utilisable par le réseau. Les polyèdres sont décomposés en voxels à l'intérieur d'un espace de $20 \times 20 \times 20$, résultant en quatre tenseurs de taille $20 \times 20 \times 20$. L'activation peut prendre des valeurs continues entre -1 et 1. Les couleurs possibles et les positions des polyèdres sont respectivement décomposés en quatre vecteurs de taille 1×3 (correspondant à l'espace RGB pour les couleurs). La présence d'une activation à une certaine position est dénotée par la valeur 1 et l'absence d'activation par la valeur -1. Le code source permettant le codage des stimuli est fourni à l'Annexe 3. Suite à l'apprentissage par le réseau de chaque attribut, les cartes topologiques résultantes, composées de 400 unités (20×20), sont assemblées en tenseurs $5 \times 20 \times 20$ pour correspondre aux cinq attributs des stimuli présentés durant la partie expérimentale. Les cartes topologiques qui forment les tenseurs ont des valeurs continues allant de -1 à 1.

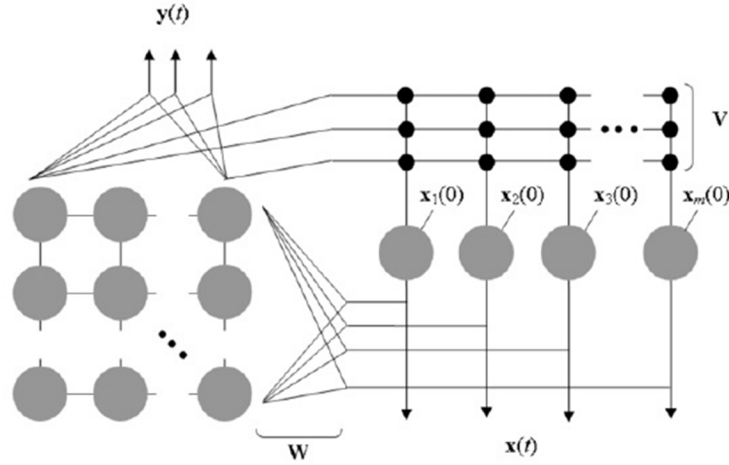
Réseau

Le réseau FEBAM-SOM été développé par Chartier, Giguère et Langlois (2009) et est écrit dans le langage de programmation du logiciel *Mathematica*. La version utilisée pour ce projet est la plus récente connue, obtenue en 2009, qui se base sur des travaux précédents (Chartier et Boukadoum 2006, Giguère, Chartier, Proulx & Lina 2007a, Giguère, Chartier, Proulx & Lina 2007b, Chartier & Giguère 2008). Il a été entièrement reprogrammé pour accélérer ses performances et pour pouvoir recevoir des stimuli tridimensionnels. Tout réseau pouvant être décrit par son architecture, fonction de transmission et règle d'apprentissage, celles-ci sont présentées sommairement ici.

Architecture

Le FEBAM-SOM est composé de deux réseaux de type Hopfield connectés tête-bêche (voir Figure 2) Chaque réseau représente une couche: la couche \mathbf{x} représentant les inputs initiaux de la phase d'apprentissage ainsi que les inputs reconstruits durant la phase de rappel. La couche \mathbf{y} est l'information comprimée qui mémorise la réduction de dimensionnalité. Deux matrices de poids sont utilisées afin de pouvoir représenter les connexions asymétriques entre les couches. La matrice de compression \mathbf{W} connecte la couche \mathbf{x} à la couche \mathbf{y} et détermine la transformation des inputs en leur version compressée. La matrice de reconstruction \mathbf{V} connecte la couche \mathbf{y} à la couche \mathbf{x} et permet la reconstruction des inputs initiaux par une combinaison non linéaire des composantes apprises.

Figure 2
Architecture du réseau



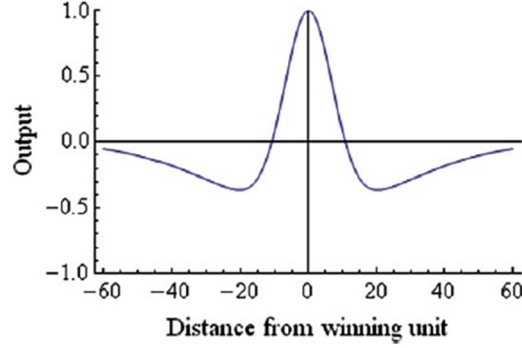
La fonction de voisinage topologique dans la couche \mathbf{y} est une fonction dite en "chapeau mexicain":

$$h_{j,i(x)}(p) = \frac{1}{2} \left(3e^{-\frac{d_{j,i(x)}^2}{\sigma^2(p)}} - e^{-\frac{d_{j,i(x)}^2}{4\sigma^2(p)}} \right)$$

$$d_{j,i(x)}^2 = \| r_j - r_i \|^2$$

où r_j représente la position sur la carte du neurone j , r_i représente la position sur la carte du neurone i , d est la distance entre deux unités, p est le numéro de l'itération en cours et σ la taille du voisinage choisi. Cette taille est constante dans la version du réseau utilisée ici.

Figure 3
Fonction de voisinage



Fonction de transmission

Les fonctions de transmission sont comme suit:

$$\forall i, \dots, N, x_i(t+1) = \begin{cases} 1, & \text{Si } Wx_i(t) > 1 \\ -1, & \text{Si } Wx_i(t) < -1 \\ (\delta + 1)Wx_i(t) - \delta(Wx_i(t))^3, & \text{Sinon} \end{cases}$$

$$\forall i, \dots, M, y_i(t+1) = \begin{cases} 1, & \text{Si } Vy_i(t) > 1 \\ -1, & \text{Si } Vy_i(t) < -1 \\ (\delta + 1)Vy_i(t) - \delta(Vy_i(t))^3, & \text{Sinon} \end{cases}$$

où N et M sont le nombre d'unité des couches \mathbf{x} et \mathbf{y} respectivement, i est l'index du vecteur, $\mathbf{x}(t)$ et $\mathbf{y}(t)$ représente les valeurs dans les couches au temps t et δ est un facteur de transmission positif qui doit être fixé à une valeur inférieure à 0.5 pour obtenir un comportement stable du réseau.

Règle d'apprentissage

Les règles d'apprentissage du FEBAM-SOM sont basées sur une association hébienne différentielle sur le temps. Les règles sont comme suit:

$$W(p + 1) = W(p) + \eta(y(0) - y(t))(x(0) + x(t))^T$$

$$V(p + 1) = V(p) + \eta(x(0) - x(t))(y(0) + y(t))^T$$

où $\mathbf{x}(0)$ et $\mathbf{y}(0)$ représente les valeurs initiale dans les couches et η est le paramètre d'apprentissage qui doit respecter la condition suivante:

$$\eta < \frac{1}{2(1 - 2\delta)\text{Max}[N, M]}, \delta \neq \frac{1}{2}$$

Déroulement de la simulation

On laisse d'abord le réseau apprendre de façon indépendante les attributs composant les stimuli, soit les polyèdres, les couleurs et les positions. Nous obtenons alors une carte topologique pour chacun des attributs. Nous construisons alors des stimuli correspondant aux stimuli présentés aux sujets humains en combinant les attributs tels que le réseau les a appris. Ensuite, les stimuli sont regroupés selon les ensembles correspondant aux tâches ayant été présentées dans la partie empirique du projet. Chacune des tâches est apprise par une nouvelle instance du réseau indépendante des deux autres. Il n'y a donc pas d'effet de l'ordre de présentation des tâches. Ici, comme dans la partie empirique, l'ordre de présentation des différents stimuli est aléatoire. Durant la phase d'apprentissage, le réseau continue à itérer jusqu'à ce qu'il atteigne un critère d'erreur ou un nombre d'itérations maximum. Le critère d'erreur s'applique sur la différence entre les stimuli initiaux et les stimuli reconstruits et était fixé à 0.01. Une fois l'apprentissage fait, le réseau passe à la phase de transfert. Durant la phase de transfert, tous les paramètres sont gardés constants avec la phase d'apprentissage, sauf le nombre d'itérations, qui était de 5000.

Résultats

Résultats empiriques:

Phase d'apprentissage

La performance des sujets a été étudiée à la fin de la phase d'apprentissage (i.e. le bloc d'essais où les participants ont atteint le critère de 90% de bonnes réponses) puis au

cours de la phase de transfert. Des tests t et des analyses de variance à mesure répétées suivies de tests *a posteriori* de Tukey ont été effectuées sauf si indication contraire.

Le nombre de blocs d'entraînement nécessaires pour que les sujets atteignent le critère varie selon la tâche ($F(2, 42) = 12.7, p < .005, R^2 = .15$). La tâche de catégorisation par air de famille (une moyenne de $M = 7.3$ blocs d'essais) a été réussie en moyenne en un nombre de blocs statistiquement équivalent à la tâche de catégorisation par règle conjonctive ($M = 8.6$ blocs), toutes deux réussies plus rapidement que la tâche de catégorisation par règle disjonctive ($M = 11.0$). Les analyses statistiques sont rapportées au Tableau IV. Les sujets qui n'ont pas appris la tâche au bout de 20 blocs d'essais ont été retirés des analyses subséquentes. Les données de 39 sujets pour la tâche de catégorisation par ressemblance familiale, 40 pour la tâche de catégorisation par règle conjonctive et 30 pour la tâche de catégorisation par règle disjonctive ont été conservées. On remarque ici très clairement que la tâche de catégorisation par règle disjonctive se démarque quant à sa difficulté subjective pour les sujets.

Tableau IV

Différences significatives dans le nombre de blocs nécessaires à l'apprentissage des tâches

	Test statistique
Différence tâche par air de famille – tâche conjonctive	$q(41) = .95, p > .05$
Différence tâche par air de famille - tâche disjonctive	$q(30) = 3.45, p < .01$
Différence tâche conjonctive - tâche disjonctive	$q(31) = 2.66, p < .05$

Comme tous les participants devaient atteindre le critère de 90% de bonnes réponses avant de passer à la phase de transfert, il y a très peu de variabilité entre les taux de bonnes réponses pour les trois tâches durant le dernier bloc d'essais. Notons aussi que les temps de réponses sont aussi équivalents entre les trois tâches. Nous avons établi comme variables contrôles l'appartenance au programme d'étude ainsi que l'ordre de passation des tâches sur la performance. Il n'y a pas de différence significative entre les sujets provenant de programme d'étude en sciences naturelles ou en génie et ceux provenant de programmes en sciences sociale, humaines ou en lettres, ni sur les temps de réponse ni sur le taux de bonnes réponses durant la phase d'apprentissage. Il n'y a pas non

plus d'influence de l'ordre de passation sur les temps de réponse et taux de bonnes réponses durant cette phase. Les moyennes se trouvent au Tableau V en Annexe I.

Nous avons évalué l'effet de la typicité, tel que représenté par la ressemblance au prototype, sur l'apprentissage des stimuli dans la tâche de catégorisation par air de famille. Les temps de réponses ne sont pas significativement différents entre les stimuli partageant quatre attributs avec le prototype de la catégorie par rapport aux stimuli en partageant trois (1329 ms vs 1491 ms). Cependant, le taux de bonne réponses moyen est meilleur pour le groupe partageant quatre attributs (97.7%) que trois (90.6%, $t(76) = 4.7$, $p < .001$, avec une taille d'effet $d = 0.76$). Notre étude trouve donc que les items ayant la plus grande typicité catégorielle sont appris plus efficacement, ce qui est en accord avec les effets de la typicité énoncés dans des études antérieures.

Nous avons aussi vérifié l'effet de la familiarité dans les tâches de catégorisation par règle logique. Rappelons que pour les deux tâches, deux niveaux de familiarité étaient mis en relief, un niveau de haute familiarité où une valeur d'attribut était présente 75% du temps et un niveau de moyenne familiarité où une valeur était présente 50% du temps. Les deux niveaux n'étant pas implémenté sur le même attribut, quatre groupes d'items peuvent être formés : aucune familiarité, moyenne familiarité, haute familiarité et haute et moyenne familiarité.

Pour la tâche de catégorisation suivant une règle conjonctive, les deux niveaux de familiarité entrent en interaction concernant les taux de bonnes réponses ($F(3, 156) = 22.0$, $p < .001$, $R^2 = .30$). Les stimuli incluant des attributs créant une haute familiarité ont un taux de bonnes réponses plus grand que les items n'incluant pas ces attributs, qu'il y ait présence d'attributs créant une moyenne familiarité ($M = 100\%$) ou non ($M = 98,3\%$, $q(156) = 0.02$, $p > .05$). Les items présentant une valeur qui était vue fréquemment sont donc catégorisés de façon plus efficace, même si cet élément de fréquence n'est pas informatif sur la règle à utiliser. Pour les items ne présentant pas d'attributs créant une haute familiarité, les items n'ayant aucun attribut créant de la familiarité ont un meilleur taux de bonnes réponses ($M = 93,3\%$) que ceux ayant des attributs créant de la moyenne familiarité ($M = 82,5\%$, $q(156) = 0.11$, $p < .001$). Ce dernier résultat est étonnant vu les

effets connus et documentés de la familiarité sur la catégorisation. Nous y reviendrons dans l'interprétation. L'observation des temps de réponse révèle qu'il n'y a pas d'interaction entre les niveaux de familiarité pour cette variable. Les items comprenant des attributs créant une haute familiarité ($M = 1321$ ms) ont des temps de réponse plus bas que les items n'ayant pas ces valeurs ($M = 1524$, $t(158) = 2.2$, $p < .05$, $d = 0.35$) alors que la moyenne familiarité n'a aucun effet sur les temps de réponse. Les taux de bonnes réponses et les temps de réponse sont présentés au Tableau VI.

Tableau VI

Taux de bonnes réponses et temps de réponse par groupes de familiarité dans les tâches de catégorisation par règle conjonctive et disjonctive

	Haute familiarité	Pas de Haute familiarité	Moyenne familiarité	Pas de Moyenne familiarité
Temps de réponse tâche conjonctive	1321 ms	1524 ms	1502 ms	1343 ms
Temps de réponse tâche disjonctive	1400 ms	1740 ms	1618 ms	1523 ms
Taux de bonnes réponses tâche disjonctive	96,1%	89,5%	96,5%	89,1%
	Haute familiarité et Moyenne familiarité	Haute familiarité seulement	Moyenne familiarité seulement	Pas de familiarité
Taux de bonnes réponses	100%	98,3%	82,5%	93,3%
tâche conjonctive				
Taux de bonnes réponses	98%	94,5%	95%	84%
tâche disjonctive				

Pour la tâche de catégorisation suivant une règle disjonctive, les deux niveaux de familiarité n'entrent pas en interaction. Les items contenant des attributs créant une haute familiarité reçoivent une réponse plus rapidement ($M = 1400$ ms, $t(107.4) = 3.5$, $p < .005$, $d = 0.63$) et plus exacte ($M = 96,1\%$, $t(81.6) = 3.2$, $p < .005$, $d = 0.58$) que ceux ne présentant pas ces attributs ($M = 1740$ ms, $M = 89,5\%$). Les items ayant des attributs créant une moyenne familiarité reçoivent des réponses plus exactes ($M = 96,5\%$, $t(118) = 3.6$, $p < .005$, $d = 0.65$) que ceux ne présentant pas ces attributs ($M = 89,1\%$). Ce niveau de familiarité n'a pas d'effet sur les temps de réponse. On peut observer que le taux de bonnes réponses entre les items présentant l'attribut créant la haute familiarité et celle présentant la moyenne

familiarité semblent à peu près égales. Lorsqu'on considère les quatre groupes simultanément, on observe que pour cette tâche, les items contenant un, l'autre ou les deux types de familiarité ont un taux de bonnes réponses équivalent et plus élevé que celui pour les items n'ayant pas d'attribut créant de la familiarité ($q(116) = .14$, $p < .001$). Ainsi, la présence de familiarité, quel que soit son niveau semble aider à la catégorisation pour la tâche disjonctive. Les temps de réponse et les taux de bonnes réponses sont présentés au Tableau VI.

Comparaison entre les deux phases

Lors du passage de la phase d'apprentissage à la phase de transfert, toutes les tâches montrent une augmentation des temps de réponse significative, ainsi qu'une diminution du taux de bonnes réponses (voir Tableau VII). Ainsi, contrairement à nos attentes, les sujets n'ont pas bien réussi à intégrer les règles de classification apprises durant la phase d'apprentissage puisqu'ils ont de la difficulté à transférer le savoir acquis. Les trois tâches ont une augmentation des temps de réponse équivalente mais une diminution des taux de bonnes réponses significativement différentes ($F(2, 56) = 11.2$, $p < .001$, $R^2 = .14$). La tâche de catégorisation par règle disjonctive est celle montrant la plus grande diminution de l'exactitude de ses réponses (25.1%), suivie par la tâche par règle conjonctive (18.7%, $q(38) = 2.0$, $p < .05$) et finalement la tâche par air de famille (13.4%, $q(38) = 2.3$, $p < .05$). La perte d'efficacité est notable et bien au-delà d'une diminution causée par la nécessité de se familiariser avec le nouvel ensemble de stimuli. Notons que les participants ont quand même partiellement intégré les règles de catégorisation car pour aucune des tâches le taux de bonnes réponses n'est descendu au niveau du hasard.

Tableau VII

Changements des temps de réponse et des taux de bonnes réponses entre le dernier bloc de la phase d'apprentissage et la phase de transfert

Tâche	Temps réponse apprentissage	Temps réponse transfert	Test statistique	Bonnes réponses training	Bonnes réponses transfert	Test statistique
Ressemblance familiale	1411 ms	1585 ms	$q(38) = 2.1$, $p < .05$	94,1%	80,3%	$q(38) = 11.4$, $p < .001$
Règle conjonctive	1403 ms	1797 ms	$q(39) = 3.3$, $p < .005$	94,1%	75,9%	$q(39) = 9.2$, $p < .001$
Règle disjonctive	1483 ms	1922 ms	$q(29) = 2.5$, $p < .05$	93,8%	68,2%	$q(29) = 14.3$, $p < .001$

Phase de transfert

Pour la phase de transfert, les temps de réponse ($F(2, 62) = 3.9, p < .05, R^2 = .11$) et les taux de bonnes réponses ($F(2, 62) = 12.3, p < .001, R^2 = .28$) diffèrent entre les tâches. Les temps de réponse pour la tâche de catégorisation par règle logique disjonctive sont significativement plus élevés et le taux de bonnes réponses plus bas que ceux pour les deux autres tâches, qui sont statistiquement équivalentes entre elles. On remarque quand même une tendance pour ces deux tâches, à savoir que la tâche de ressemblance familiale a un taux de bonnes réponses plus élevé que la tâche de catégorisation par règle conjonctive. Ces résultats nous donnent encore ici une indication du niveau de difficulté clairement supérieur de la tâche disjonctive. Le tableau VIII présente les moyennes et les tests statistiques et un graphique présentant les taux de bonnes réponses lors des deux phases est présenté à la Figure 2.

Tableau VIII

Temps de réponse (tests présentés dans la matrice triangulaire du haut) et taux de bonnes réponses (matrice triangulaire du bas) durant la phase de transfert

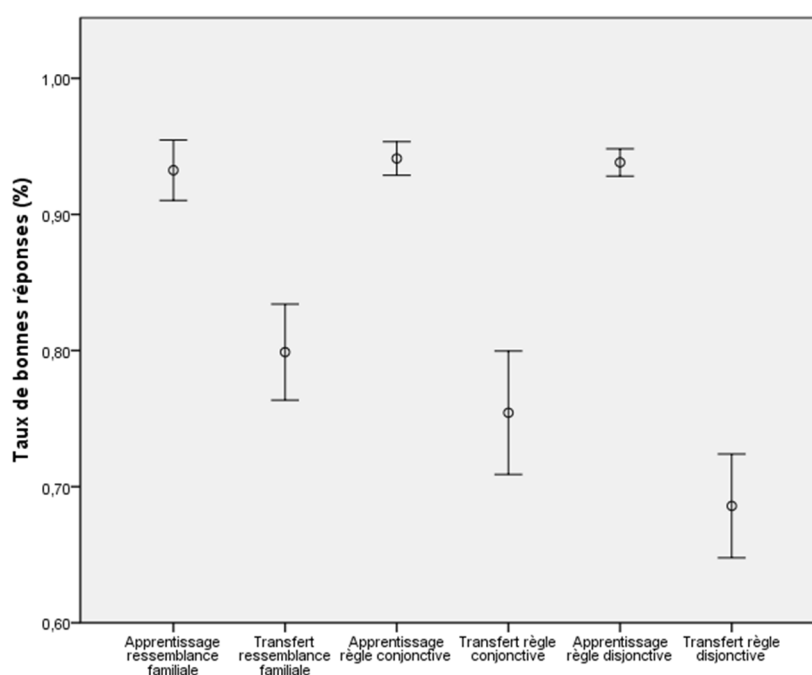
		Ressemblance familiale	Règle conjonctive	Règle disjonctive
		1585 ms	1797 ms	1922 ms
Ressemblance familiale	80,3%		$t(38) = 1.3, p > .05$	$t(31) = 2.2, p < .05$
Règle conjonctive	75,9%	$t(38) = 1.9, p > .05$		$t(32) = 2.6, p < .05$
Règle disjonctive	68,2%	$t(31) = 6.1, p < .001$	$t(32) = 2.8, p < .01$	

Les deux variables pour lesquelles nous avons contrôlé, soit le domaine d'étude et l'ordre de passation des tâches, ont eu un effet durant la phase de transfert. Les étudiants en sciences humaines sont plus rapides sans être plus efficaces durant la tâche de catégorisation par règle conjonctive (1344 ms vs 1886 ms, $t(30) = 4.6, p < .05, d = .79$). Par contre, durant la tâche de catégorisation par règle disjonctive, les étudiants en sciences naturelles se sont montrés plus efficaces (71,7% vs 64,4%, $t(30) = 5.0, p < .05, d = .72$). Tous les autres temps de réponse et taux de bonnes réponses sont équivalents. Bien que ces résultats soit net statistiquement, peu de choses peuvent être généralisées à partir de ceux-ci. En effet, si on souscrit à l'idée que les étudiants en science naturelles ont une plus grande facilité avec la logique, ils devraient aussi avoir mieux réussi la tâche de catégorisation par règle conjonctive, or ce résultat n'a pas été observé.

Concernant l'ordre de passation des tâches, les sujets ayant effectué la tâche de catégorisation par règle conjonctive après la tâche par ressemblance familiale sont plus rapides (1446 ms vs 2225 ms, $t(20) = 2.3$, $p < .05$, $d = .73$) et plus exacts (79,6% vs 71, 5%, $t(38) = 2.2$, $p < .05$, $d = .70$) que ceux l'ayant effectué après la tâche disjonctive. Finalement, les sujets ayant effectué la tâche de catégorisation par règle disjonctive en dernier sont plus rapides à celle-ci que ceux l'ayant faite en premier (1502 ms vs 2428 ms, $t(16) = 2.8$, $p < .05$, $d = .97$). Cet effet est facilement explicable par un effet de pratique, qui est d'autant plus visible que cette tâche est la plus difficile à apprendre.

Figure 4

Taux de bonnes réponses durant la phase d'apprentissage et la phase de transfert



Rappelons que durant la phase de transfert de la tâche de catégorisation par intégration d'information, deux groupes d'items étaient visionnés. Le premier groupe contient des items introduisant une valeur d'attribut jamais vue auparavant alors que le deuxième groupe est composé de nouveaux items mais ne contenant que des valeurs déjà vues. Nous avons trouvé que les deux groupes d'items ont des taux de bonnes réponses différents ($F(5, 228) = 10.7$, $p < .001$, $R^2 = .19$) mais des temps de réponse équivalents. Les items contenant la valeur nouvelle et ayant quatre ou trois attributs en commun avec le prototype obtiennent des réponses plus exactes que ceux contenant deux attributs (93,6%

et 86,2% vs 71,5%, $q(5, 228) = .22$, $p < .001$), ce qui représente une répétition de l'effet de typicité trouvé durant la phase d'apprentissage. Il semble par contre que passé un certain niveau de typicité, il n'y a pas de différence quant à l'efficacité de catégorisation. Il n'y a pas de différence quant à l'exactitude des items nouveaux contenant 4, 3 ou 2 éléments d'un item déjà vu. Par contre, pour les items ayant quatre attributs appartenant au prototype, ceux qui introduisent une valeur nouvelle sont mieux réussis que ceux n'utilisant que des attributs déjà vus (93,6% vs 77,9%, $q(2,228) = .16$, $p < .001$). Ce résultat est étonnant et sera abordé plus avant dans la discussion.

Pour la tâche conjonctive, les stimuli de la phase de transfert ayant un attribut créant de la familiarité sont les mieux réussis (83,1% vs 68,8%, $t(70) = 3.9$, $p < .001$, $d = .66$). De ces items, ceux appartenant à la catégorie A ont un taux de bonne réponse plus élevé que ceux n'y appartenant pas (84,6% vs 66,3%, $t(69) = 2,8$, $p < .01$, $d = .48$). En effet, il est logique que la familiarité permette de mieux catégoriser les items associés à une catégorie plutôt que ceux au contraire n'ayant pas de caractéristiques les définissant. Il n'y a aucune différence pour les temps de réponse.

Rappelons que durant la phase de transfert de la tâche de catégorisation par règle disjonctive, trois groupes d'items étaient visionnés. Le premier groupe contient des items introduisant une valeur d'attribut jamais vue auparavant alors que les deuxième et troisième groupes sont respectivement composés de nouveaux items ne contenant que des valeurs déjà vues partageant quatre et trois attributs avec un item déjà vu. Les différents groupes de stimuli ont des taux de bonnes réponses différents ($F(2, 96) = 41.0$, $p < .001$, $R^2 = .46$) *mais des temps de réponse équivalents. Les éléments incorporant une valeur nouvelle obtiennent un taux de bonne réponse plus élevé que les items n'incorporant pas cette valeur mais partageant quatre attributs en commun avec un item déjà vu* ($q(2, 96) = 0.14$, $p < .001$), *eux même ayant une exactitude supérieure aux éléments partageant trois attributs avec un item précédemment vu* (79,4% vs 65,2% vs 49,0%, $q(2,96) = 0.16$, $p < .001$). *Il est étonnant que des stimuli intégrant une valeur nouvelle soient mieux réussis que des stimuli ne contenant que des valeurs connues. Par contre, si on contrôle pour la familiarité (qui a un effet tel que montré durant la phase d'apprentissage), on voit que les stimuli ayant une haute familiarité sont équivalents, qu'ils contiennent la valeur nouvelle ou la moyenne familiarité ou non. Ces stimuli ont un meilleur taux de bonnes réponses que les stimuli à moyenne familiarité (indépendamment de la nouveauté), eux même mieux*

réussis que les stimuli ayant aucune familiarité mais n'incorporant pas la valeur nouvelle (voir Tableau IX). On remarque le résultat étonnant que les éléments sans familiarité mais nouveaux sont réussis au même niveau que ceux ayant une haute familiarité. Par contre, un seul stimulus correspond à ce descriptif et il est donc possible que ce résultat soit un artefact.

Tableau IX

Taux de bonnes réponses pour les différents niveaux de familiarité durant la phase de transfert de la tâche de catégorisations par règle disjonctive

Groupe	Haute fam.	Moyenne fam	Groupe	Moy.	Inter-groupes	Différence significatives
Valeurs jamais rencontrées	Oui	Oui	1	82.1%	1-5	$q(77) = .18, p < .005$
		Non	2	85.5%	1-15	$q(77) = .28, p < .001$
	Non	Oui	3	63.4%	1-16	$q(77) = .48, p < .001$
		Non	4	93.8%	2-5	$q(77) = .22, p < .001$
Valeurs déjà connues	Oui	Non	5	80.9%	2-15	$q(77) = .33, p < .001$
	Non	Oui	6	53.0%	2-16	$q(77) = .53, p < .001$
		Non	7	33.0%	5-6	$q(76) = .30, p < .001$
					5-12	$q(77) = .18, p < .005$
					5-16	$q(77) = .30, p < .001$
					6-15	$q(76) = .41, p < .001$
					6-16	$q(76) = .61, p < .001$
					12-15	$q(77) = .28, p < .001$
					12-16	$q(77) = .48, p < .001$
					15-16	$q(77) = .20, p < .001$

Résultats de la modélisation

La première étape de la modélisation du processus d'apprentissage de catégories par le réseau était une phase d'apprentissage des différentes valeurs possibles pour les attributs. Pour ce faire, nous avons fait apprendre chacun des attributs séparément au réseau. Les paramètres de l'apprentissage se trouvent au Tableau X. Nous avons réussi à obtenir une erreur (calculée comme la somme des différences entre le tenseur du stimulus initial et le tenseur du stimulus reconstruit au carré, le SSE) très faible (sous .002) pour les trois types d'attributs. On peut donc penser que les cartes topologiques de chacune des valeurs possibles pour ces attributs sont des représentations stables de ceux-ci.

Tableau X
Paramètres d'apprentissage des attributs uniques

Attribut	Nombre d'itérations	Sigma	Delta	Taille de la carte	Nombre gagnants	Taille des stimuli	Erreur
Forme	10000	4	.1	10 × 10	20	20 × 20 × 20	.0015
Couleur	110	4	.1	10 × 10	20	3 × 1	.001
Position	246	4	.1	10 × 10	20	3 × 1	.001

Ensuite, nous avons simulé la tâche telle que les participants l'ont effectuée. Étant donné qu'il faut plusieurs heures (environ 10) par simulation, il n'est pas possible de répéter les simulations un grand nombre de fois. Nous avons opté pour 4 simulations, en variant légèrement les paramètres d'apprentissage. Le Tableau XI présente ces paramètres.

Dans aucune des simulations le réseau a réussi à atteindre le critère d'erreur minimal que nous avons fixé à .01. Chacune des simulations atteint donc son critère d'arrêt, soit 50 000 itérations (sauf pour la 4e tâche, où nous avons arrêté le réseau après 5 000 itérations car l'erreur avait atteint une asymptote). Pour chaque simulation, la figure présentée est celle provenant de la simulation présentant les résultats les plus clairs.

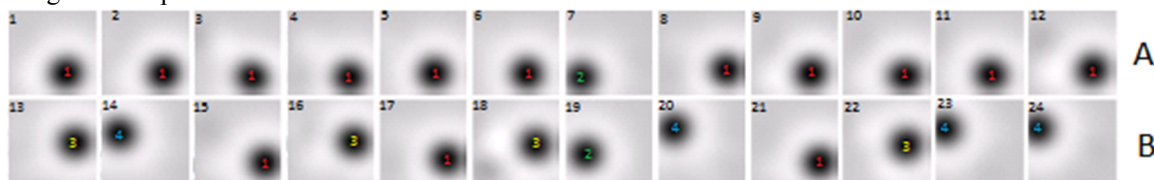
Tableau XI
Paramètres des trois tâches de catégorisation effectuées par le réseau. En gras se trouve le paramètre modifié. La tâche 1 est la tâche par air de famille, la tâche 2 est la tâche conjonctive et la tâche 3 est la tâche disjonctive.

Essai	Tâche	Sigma	Delta	Taille carte	Nombre gagnants	Erreur
1	1	4	.001	20 × 20	20	13.6
	2					4.05
	3					4.53
2	1	4	.001	16 × 16	20	9.24
	2					4.69
	3					6.89
3	1	4	.01	20 × 20	25	5.93
	2					3.33
	3					10.6
4	1	8	.001	20 × 20	20	9.97
	2					6.95
	3					9.40

On trouve que pour la tâche de catégorisation par air de famille, le réseau fait significativement plus d'erreur à la fin de l'apprentissage ($M = 9.69$) que pour la tâche de catégorisation par règle conjonctive ($M = 4.76$, $q(2,9) = 4.9$, $p < .05$). Cette dernière est d'ailleurs celle ayant été apprise avec moins d'erreur finale pour les quatre simulations. Aucune des deux tâches ne diffère significativement de la tâche de catégorisation par règle disjonctive ($M = 7.85$).

Nous avons examiné les résultats pour chaque tâche séparément. Pour la tâche de catégorisation par air de famille, nous nous attendions à ce que le réseau crée deux attracteurs, soit un par catégorie, avec possiblement un voisinage étendu reflétant la ressemblance plus ou moins grande au prototype. Sur les cartes topologiques représentant l'activation correspondant à la présentation de chaque stimulus, cela se verrait par une activation de la carte au même endroit pour les éléments partageant le même attracteur. Ce qui peut être observé sur les cartes topologiques (Figure 5) est tout autre. La ligne supérieure correspond aux réponses du réseau lorsque des stimuli de la catégorie A sont présentés alors que la ligne du bas correspond aux stimuli de la catégorie B. Chaque colonne correspond à un stimulus différent, identifié par le numéro dans le coin supérieur gauche (voir Tableau I). Les attracteurs sont identifiés à l'aide de numéros. On peut voir que le réseau a de façon générale bien classé ensemble les éléments appartenant à la catégorie A car les activations sont toutes (sauf une) situées dans le coin inférieur droit des cartes. Les stimuli appartenant à la catégorie B n'ont par contre pas de position systématique. On peut observer que le tiers de ceux-ci sont dans le coin supérieur gauche, un autre tiers au centre droit, trois à la même position que la catégorie A et un est seul dans le coin inférieur gauche. On ne peut donc pas parler d'attracteur unique pour la catégorie B.

Figure 5
Carte topologique des activations du réseau durant la phase d'apprentissage pour la tâche de catégorisation par air de famille.



Nous avons aussi vérifié l'effet de la typicité sur les performances du réseau. Nous avons fait deux simulations supplémentaires dans lesquelles le réseau devait apprendre séparément les stimuli partageant quatre attributs avec le prototype et ceux partageant trois attributs avec celui-ci. Nous avons obtenu un résultat inverse à celui obtenu pour les sujets humains en ce qui concerne l'effet de la typicité. Le réseau a appris les items ayant trois éléments en commun avec le prototype plus rapidement et avec moins d'erreur que ceux ayant quatre éléments communs avec celui-ci dans les deux simulations effectuées par le réseau. Les résultats se trouvent au Tableau XII. Nous reviendrons à ce résultat dans la discussion.

Tableau XII

Résultats de l'apprentissage de la tâche de catégorisation par air de famille lorsque les stimuli de typicité 4 sont appris séparément des autres stimuli.

Essai	Typicité(nb de valeurs)	Itérations	δ	Taille de la carte	Nombre de gagnants	Erreur	Temps (sec)
1	4	50000	.001	20×20	20	2.62	23226
	3					0.01	18772
2	4		.01		25	0.82	22705
	3					0.01	19779

Durant la phase de transfert, les nouveaux items appartenant à la catégorie A ont été aussi bien classé pour les item ayant une valeur nouvelle jamais encore vue (Figure 6) que pour ceux n'utilisant que des valeurs déjà vues par le réseau (Figure 7). On peut voir dans la Figure 6 que la majorité des cartes des stimuli appartenant à la catégorie A sont encore ici situés dans le coin inférieur droit (8 sur 12) ou généralement à droite (2 sur 12). Pour les items n'incluant pas de nouveauté, 9 sur 12 se trouvent à l'emplacement de l'attracteur créé durant la phase d'apprentissage alors que 2 se trouvent généralement à droite. Par contre, pour ce qui est de la catégorie B, les items avec une valeur nouvelle semblent avoir une classification plus constante (du côté gauche, 9 sur 12) que ceux n'incluant que des valeurs connues (3 sur 12).

Figure 6

Carte topologique des activations du réseau pour les stimuli avec une valeur nouvelle durant la phase de transfert pour la tâche de catégorisation par air de famille (même format qu'à la Figure 3).

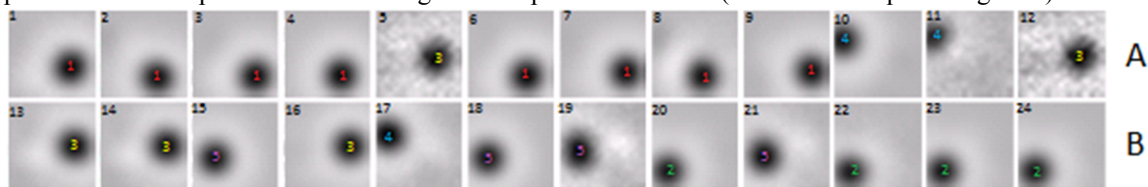
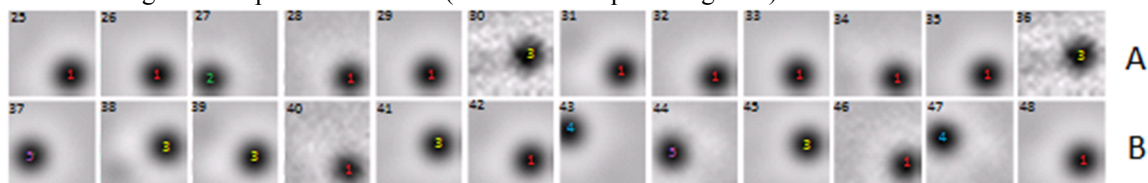


Figure 7

Carte topologique des activations du réseau pour les stimuli durant la phase de transfert pour la tâche de catégorisation par air de famille (même format qu'à la Figure 3).



Pour la tâche de catégorisation par règle conjonctive, nous nous attendions à trouver un attracteur bien défini pour la catégorie A et déterminé par la valeur des deux attributs discriminants 1 et 2. Nous ne prédisions pas de création d'attracteur systématique pour la catégorie Non-A. On peut voir à la Figure 8 l'arrangement des activations pour chacun des stimuli. Les stimuli appartenant à la catégorie A sont en effet arrangés de façon beaucoup plus structurée que les stimuli appartenant à la catégorie non-A. Par contre, il n'y a pas eu de création d'attracteur unique pour cette catégorie. Les stimuli sont groupés trois par trois et chaque groupe a son attracteur (l'attracteur 1 revenant plus souvent). De façon générale, les stimuli ont été groupés à l'intérieur de la catégorie par leur ressemblance en utilisant les attributs non discriminants, principalement l'attribut 4 avec un petit apport de l'attribut 3. Vous pouvez vous référer au Tableau II pour voir les valeurs des attributs. Sur la Figure 8, le groupement par cet attribut non discriminant se démontre par le positionnement exactement identique pour les stimuli 1 à 3 et 10 à 12, 4 à 6, et 7 à 9 respectivement, qui appartiennent tous à la catégorie A. Lorsqu'on observe les stimuli 13 à 24 appartenant à la catégorie Non-A, ces groupements par trois ne sont pas présents de façon aussi systématique, mais encore ici, les attributs 4 et 3 sont ceux qui semblent avoir été utilisés par le réseau pour sa catégorisation. On peut en conclure que les attributs notant l'appartenance au groupe n'ont

donc pas été utilisés prioritairement mais on plutôt eu un effet stabilisateur sur la façon dont les stimuli ont été groupés. On peut aussi interpréter les résultats obtenus en considérant que le réseau a catégorisé encore ici par air de famille et les stimuli appartenant à A partageaient tous au moins 2 attributs, en plus de ceux utilisés dans les groupements, ce qui a permis de créer des attracteurs plus stables que pour la catégorie non-A.

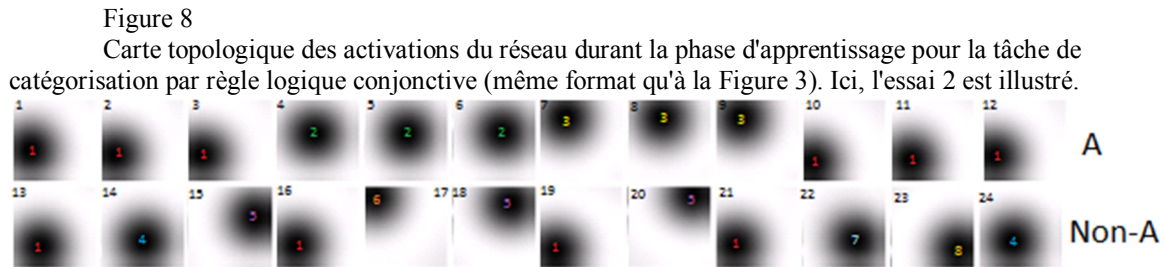
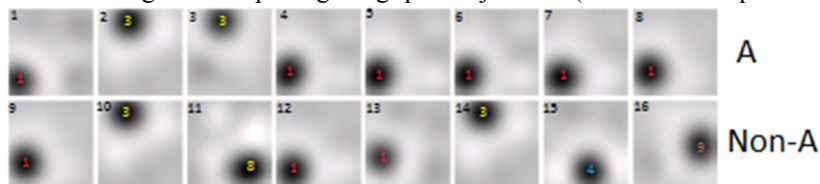


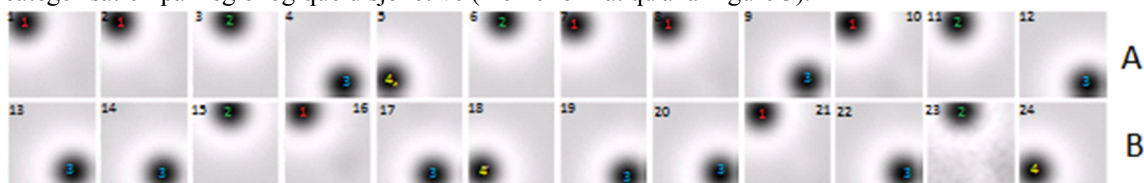
Figure 9
Carte topologique des activations du réseau pour les stimuli durant la phase de transfert pour la tâche de catégorisation par règle logique conjonctive (même format qu'à la Figure 3).



Durant la phase de transfert, l'ajout d'une valeur nouvelle a forcé le réseau à utiliser un autre attribut que le 4^e pour regrouper les stimuli puisque celui-ci est celui contenant la valeur nouvelle. Effectivement, aucun poids de connexion n'est associé à la nouvelle valeur. Comme cet attribut est celui qui a été utilisé prioritairement durant la phase d'apprentissage, cela a pour effet d'unifier les stimuli A en une catégorie unique utilisant majoritairement les attributs déterminants pour la catégorie (Figure 9). Ceci est observable dans la figure par la consolidation des activations autour de l'attracteur situé dans le coin inférieur gauche. Il n'y a donc que deux attracteurs encore présents des trois initialement présents durant la phase d'apprentissage. Remarquons que les stimuli de la catégorie non-A ne sont pas plus unifiés que durant la phase d'apprentissage car les attributs déterminants pour la catégorie A ne contenait déjà aucune information pertinente pour leur catégorisation.

Finalement, pour la tâche de catégorisation par règle logique disjonctive, nous nous attendions à trouver quatre attracteurs, selon chacun des agencements possibles des deux attributs discriminants 1 et 2 (10xxx et 01xxx pour les A; 11xxx et 00xxx pour les B). Quatre attracteurs ont en effet été trouvés, mais le réseau semble avoir classé les stimuli d'abord selon leur valeur à l'attribut 4, puis selon leur valeur à l'attribut 3 (xxx2x pour l'attracteur marqué 1, xxx1x pour le 2, xxx0x pour le 4), plutôt qu'en utilisant les attributs discriminants. Les activations sont présentées à la Figure 10. La catégorie A présente pour 8 de 12 stimuli des attracteurs dans la partie supérieure de la carte alors que la catégorie B présente le même ratio d'activations dans la partie inférieure. Encore ici, le réseau a classé les stimuli sans donner priorité aux attributs définissant la règle mais bien par ressemblance familiale.

Figure 10
Carte topologique des activations du réseau durant la phase d'apprentissage pour la tâche de catégorisation par règle logique disjonctive (même format qu'à la Figure 3).



Pour la phase de transfert, les stimuli incluant une nouvelle valeur sont séparés en deux patrons d'attracteurs différents qui peuvent être observés à la Figure 11. Pour chacune des deux catégories, les attracteurs des six derniers stimuli sont le miroir des six premiers appartenant à la catégorie. Comme la valeur nouvelle était présente dans l'attribut 5, le réseau a pu utiliser ses poids de connections principaux pour cet attribut. On peut aussi voir que les items se sont séparés de façon plus claire en trois selon l'attribut 4. Nous avons maintenant 10 items dans l'attracteur 1, 8 dans le deuxième et 6 dans le troisième. (Figure 11). Pour les stimuli nouveaux mais n'intégrant pas de valeur nouvelle (Figure 12), l'utilisation du 4^e attribut pour la catégorisation n'est pas aussi clair. En effet l'attracteur 1 contient des items suivant le patron xxxx1, l'attracteur 2 xxx1x, l'attracteur 3 xx1xx et l'attracteur 4 xxx0x.

Figure 11

Carte topologique des activations du réseau pour les stimuli avec une valeur nouvelle durant la phase de transfert pour la tâche de catégorisation par règle logique disjonctive (même format qu'à la Figure 3).

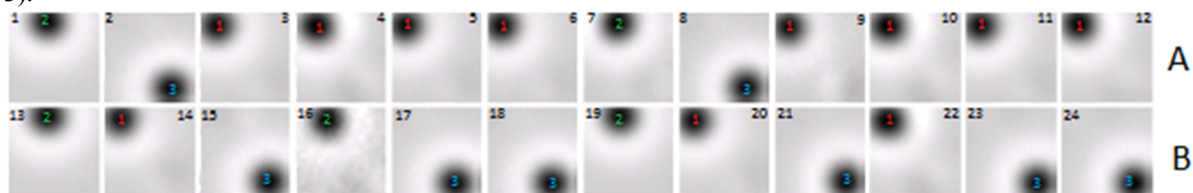
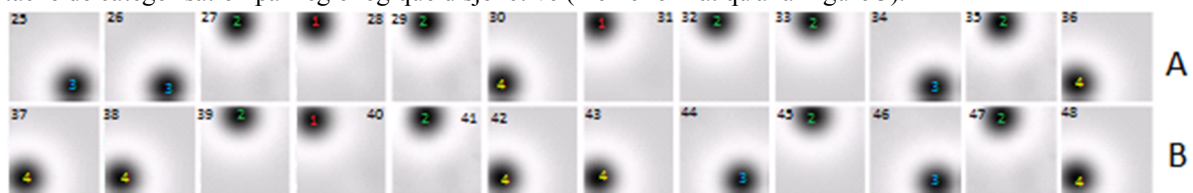


Figure 12

Carte topologique des activations du réseau pour les stimuli durant la phase de transfert pour la tâche de catégorisation par règle logique disjonctive (même format qu'à la Figure 3).



Interprétation

Résultats empiriques

Le nombre de blocs nécessaire pour apprendre chacune des tâches est cohérent avec ce qui a été obtenu par des études antérieures. La tâche de catégorisation par règle disjonctive est jugée la plus difficile par Feldman (2000) et c'est en effet la tâche ayant pris le plus de blocs d'apprentissage pour être réussie. De plus, lors du passage de la phase d'apprentissage à la phase de transfert, cette tâche est celle où on observe la plus grande baisse du taux de bonnes réponses. C'est donc la règle de catégorisation qui a été la moins bien apprise puisque les sujets parviennent difficilement à la généraliser à de nouveaux items. De façon plus intéressante, nous avons trouvé que la tâche par air de famille est celle qui a été la mieux apprise, le taux de bonnes réponse lors du passage entre la phase d'apprentissage et la phase de transfert étant le plus bas des trois tâches. De plus, bien que la différence ne soit pas statistiquement significative, la tâche de catégorisation par air de famille est aussi celle apprise le plus rapidement et celle étant la mieux généralisée aux nouveaux items lors de la phase de transfert. Nous avancerons en conséquence qu'avec les résultats en notre possession, la tâche de catégorisation par air

de famille semble la plus facile à apprendre que la tâche de catégorisation par règle conjonctive.

La typicité a un effet qui est bien documenté dans la littérature (Rosch & Mervis, 1975, Ashcraft 1978). En effet, plus un item se rapproche du prototype de la catégorie, d'un grand nombre d'exemplaires ou du centre de l'espace délimité par la frontière catégorielle et plus cet item recevra une réponse de catégorisation qui est rapide et sans erreur. Ce même effet a été observé dans la tâche de catégorisation par air de famille où les items se rapprochant le plus du prototype de la catégorie sont ceux qui ont été classés le plus rapidement et avec la plus grande exactitude et ce que ce soit durant la phase d'apprentissage ou celle de transfert.

L'influence de la familiarité sur la catégorisation est connu depuis l'étude de Malt et Smith (1982) mais a été retrouvé de multiples fois depuis dans des domaines de catégorisation différents (Kleine & Kernan 1992; Hannah & Brooks, 2009). Les items familiers sont classés plus rapidement et avec moins d'erreur et leur présence peut aider au rappel d'information apprise simultanément avec ceux-ci. Dans notre étude, nous avons observé un effet de la familiarité dans la tâche conjonctive, tant au niveau de la phase d'apprentissage que de la phase de transfert. Les items présentant une haute familiarité reçoivent des réponses plus rapides et plus exactes que les autres items. Il est par contre étonnant de trouver que les items présentant une familiarité moyenne ne sont pas mieux réussis que ceux ne présentant aucune familiarité. Nous n'avons pas d'explication pour ce résultat. Pour la tâche disjonctive, nous avons observé des effets équivalents pour la haute et la moyenne familiarité sur le taux de bonnes réponses. Ces résultats suggèrent qu'il y aurait un niveau de difficulté de la tâche de catégorisation par règle logique tel qu'il n'y a aucun gain d'efficacité à utiliser la familiarité pour faire le classement, la règle étant assez aisée pour pouvoir être utilisée directement. Plus la règle deviendrait complexe et plus les sujets devraient trouver des moyens externes à la règle pour arriver à classer les items. La règle conjonctive, relativement facile, ne ferait usage que de la haute familiarité, les niveaux de familiarité plus bas n'étant pas nécessaires à une réussite de la catégorisation alors que la règle disjonctive, plus difficile, mobiliserait tous les indices de familiarité disponibles.

Problèmes méthodologiques

Quelques problèmes méthodologiques ont pu diminuer la portée de nos résultats. Premièrement, il existe deux problèmes avec les stimuli choisis. D'abord, comme les stimuli étaient un assemblage de deux polyèdres, les attributs de formes et de couleur observés pouvaient facilement être confondus entre un polyèdre et l'autre, ce qui rendait les tâches plus difficiles. La répétition des attributs rendait donc les attributs non unitaires car pour chaque attribut, le sujet devait aussi se souvenir du polyèdre auquel il rapportait. Aussi, comme chaque ensemble de stimuli était composé d'un nombre relativement faible d'items (24), les sujets ont pu utiliser la mémorisation comme stratégie alternative pour donner la bonne réponse lorsque l'apprentissage de la règle semblait plus difficile. Cela a eu pour effet de faire en sorte que lors de la phase de transfert, les taux de bonnes réponses ont chuté drastiquement, ce qui ne se serait pas produit si les sujets avaient véritablement appris les règles de catégorisation. Une solution pour pouvoir utiliser cet ensemble de stimuli serait d'augmenter le nombre d'items dans chaque tâche afin de faire de la mémorisation une stratégie inefficace et de changer les attributs redondants de stimuli pour d'autres (présenter les polyèdres sur un fond de couleur et présenter des polyèdres de différentes tailles, par exemple). Il y avait une possibilité que l'existence de seulement deux catégories allait faire en sorte que les sujets n'apprennent que la catégorie A et considèrent tout le reste comme "Non-A" pour les tâches disjonctive et d'intégration d'information. Ceci ne s'est pas concrétisé car aucune tâche ne montre une différence selon la catégorie, que ce soit pour la phase d'apprentissage ou de transfert.

Finalement, le niveau de difficulté de la tâche disjonctive était trop grand. Nous avons utilisé l'information provenant de la littérature comme quoi il est possible pour des sujets d'apprendre une règle disjonctive à 2 attributs mais ce que nous avons observé dans cette expérience est tout autre. En effet, cette tâche a été considérée comme très difficile à apprendre par tous les sujets, même ceux ayant atteint le critère de performance de 90%, et plusieurs ont rapporté avoir appris les items par cœur. Cela cause un problème pour l'interprétation des résultats obtenu car il est impossible de départager les gens ayant appris par cœur les items et ceux ayant appris la règle, soit le phénomène que nous voulions réellement observer. De plus, puisqu'un nombre très inégal de participants ont réussi à apprendre chacune des tâches à l'intérieur de notre critère de nombre de

répétitions maximal (20 blocs) à cause de la tâche disjonctive, il est possible que les gens ayant réussis sont qualitativement différents de ceux ayant échoué. De concert avec les changements proposés plus haut, le fait d'enlever le critère d'arrêt sur le nombre de répétitions pourrait permettre à tous d'apprendre la tâche afin de ne pas perdre trop de participants.

Résultats de la modélisation

L'utilisation du réseau FEBAM-SOM comme modèle du processus de catégorisation humain a frappé plusieurs écueils. Nous présenterons plus avant les problèmes rencontrés avec le réseau dans la section suivante. Nous allons d'abord présenter ce qui est ressorti des simulations. Alors que dans les résultats empiriques, la tâche disjonctive s'est démarquée comme étant la plus difficile, pour le réseau les résultats sont contradictoires. D'une part, la tâche de catégorisation par air de famille est celle qui est complétée avec le plus d'erreur finale, alors que les deux autres tâches ne sont pas significativement différentes. Par contre, le réseau a développé des attracteurs plus conformes aux résultats humains pour cette tâche que pour les deux autres. Cette réussite n'est par contre que partielle, la création de l'attracteur unique de la catégorie A s'est faite de façon efficace, mais la catégorie B n'a pas d'attracteur unique. On s'explique mal ce résultat, la catégorie B étant exactement identique à la catégorie A dans sa structure, avec pour seul changement une substitution des valeurs "1" en valeur "2".

La typicité a été aussi modélisée pour la tâche de catégorisation par air de famille, ou les attributs partageant 3 et 4 attributs avec le prototype ont été appris séparément. Étonnamment, les deux simulations faites ont montré un avantage tant pour la rapidité d'apprentissage que pour le niveau d'erreur dans le cas de l'apprentissage pour des items à 3 attributs partagés avec le prototype, ce qui est l'inverse du résultat obtenu chez les participants. L'explication que nous avançons est que contrairement aux humains pour qui une plus petite variabilité intra-catégorielle permet de reconnaître les items comme appartenant à la même catégorie, le réseau utilise ces petites différences pour différencier les items semblables et ainsi éviter de leur assigner le même attracteur et donc perdre leur information individuelle. Rappelons-nous que l'erreur est calculée sur la différence entre l'item reconstruit et le stimulus initial. Un plus haut taux d'erreur pourrait donc indiquer

une représentation plus compacte de plusieurs items en un même attracteur. Il est donc probable que ce réseau ne cherche pas à rassembler les stimuli, mais plutôt à les distinguer (en minimisant les ressemblances inter-catégorie plutôt qu'en maximisant les ressemblances intra-catégorie).

Au cours de l'apprentissage des trois tâches, il est ressorti de façon consistante mais aussi inattendue que le réseau crée ses attracteurs en considérant tous les attributs, bien qu'à des niveaux très différents. Il était attendu pour les deux tâches de catégorisation par règle logique que puisque l'information contenue dans les deux premiers attributs était discriminante, les poids de connections reliés à ceux-ci seraient les plus forts et dirigeraient la classification. Les résultats montrent plutôt que le réseau a donné priorité aux attributs 4 et 3. Le réseau a donc choisi les attributs avec une variabilité moyenne. En effet, les attributs discriminants étaient fixes, l'attribut 3 présentait une répartition inégale des valeurs due à l'implémentation de la haute familiarité, l'attribut 4 présentait une répartition inégale des valeurs due à l'implémentation de la moyenne familiarité et l'attribut 5 présentait des valeurs équiprobables. Nous reviendrons dans la section suivante sur une tentative d'explication de ce choix.

Pour la tâche de catégorisation par règle conjonctive, les items de la catégorie A sont très clairement classés selon l'attribut 4. Lorsque l'attribut nouveau a remplacé celui-ci durant la phase de transfert, le réseau a dû utiliser les autres poids de connexions moins importants, ce qui donne un attracteur beaucoup plus constant dans le coin inférieur gauche (Figure 7, 6 items sur 8). Tous les items n'appartenant pas à la catégorie A n'ont pas de structure qui les unissent, ce qui explique le très grand nombre d'attracteurs différents créés par le réseau.

Finalement, pour la tâche de catégorisation par règle disjonctive, nous avons trouvé quatre attracteurs, bien que ce ne fussent pas ceux attendus. Trois des attracteurs représentent les valeurs possibles de l'attribut 4 lors de la phase d'apprentissage alors que le dernier attracteur, contenant moins d'items, ne semble pas avoir d'attribution systématique, un peu comme la catégorie non-A de la tâche conjonctive. Durant la phase

de transfert, l'insertion d'un nouvel attribut consolide les attracteurs autour de l'attribut 4, n'activant que les 3 attracteurs reliés aux valeurs de cet attribut.

Modifications à apporter au réseau

Plusieurs améliorations sont nécessaires si on veut que le réseau FEBAM-SOM puisse être utilisé pour modéliser le processus de catégorisation. Rétrospectivement, il était probablement mal avisé de penser qu'un réseau unique pourrait représenter fidèlement deux processus catégoriels fondamentalement différents. Les auteurs du COVIS (Ashby et al., 1998) ont sans doute réalisé ce problème puisqu'ils ont développé un modèle ayant deux systèmes de traitement. Le réseau FEBAM-SOM en sa forme actuelle a de bonnes capacités d'adaptation mais celles-ci ne suffisent pas pour s'ajuster aux données empiriques.

Commençons d'abord par présenter un problème qui relève de la méthodologie employée plutôt que du réseau lui-même. En effet, dans notre choix de vouloir un réseau parfaitement autoapprenant, nous avons négligé de faire correspondre les indices disponibles pour les sujets humains et pour le réseau. Alors que nous fournissions une rétroaction aux sujets afin de les diriger dans leur catégorisation, nous exigeons du réseau qu'il apprenne les catégories de façon entièrement indépendantes, ce qui s'avère impossible pour les catégories régies par des règles logiques arbitraires. L'étude de Chartier, Boukadoum & Amiri (2009) a d'ailleurs montré qu'il est possible d'incorporer une telle rétroaction à une mémoire associative bidirectionnelle.

Nous devons aussi cerner les problèmes actuels du réseau pour compléter les tâches de catégorisation. Premièrement, le réseau semble catégoriser par ressemblance familiale, et ce indépendamment de la structure catégorielle (ressemblance familiale ou règle logique). Lors d'une catégorisation, le réseau prend en considération tous les attributs en leur accordant une pondération qui est indépendante de leur utilité pour la tâche tel que nous voulions qu'elle soit effectuée. Dans la tâche de catégorisation par air de famille, donner une pondération égale à chaque attribut est une stratégie valide qui correspond bien aux résultats humains puisque tous les attributs ont la même importance. Par contre, pour les deux tâches de catégorisation par règle logique, le réseau n'arrive pas

à se restreindre aux attributs représentant les invariants de la catégorie. Dans ces tâches, le réseau accorde certes une pondération supérieure à certains attributs. Or, les attributs choisis ne correspondent pas aux attributs invariants qui contiennent les valeurs définissant la catégorie mais bien des attributs ayant une variabilité moyenne sans être aléatoire. Rappelons que pour effectuer une catégorisation correcte, le réseau doit accorder un poids égal entre eux, mais supérieur aux trois autres attributs, aux attributs 1 et 2. De plus, il doit utiliser deux attributs simultanément et en priorité alors que la catégorisation observée n'en utilise qu'un de façon prioritaire et les autres de façon subordonnée. Il est difficile d'expliquer pourquoi le réseau s'est principalement basé sur ce 4^e attribut (trois valeurs de l'attribut présentées) alors que le 5^e attribut (quatre valeurs présentées) présente plus de variabilité. Dans la catégorisation effectuée par le réseau, on voit que les attributs invariants ne sont pas utilisés dans la création des attracteurs et ne dirigent pas leur positionnement sur la carte topologique mais servent plutôt d'éléments stabilisants, tel qu'observé le plus clairement dans la tâche de catégorisation par règle conjonctive. En effet, dans cette tâche, on remarque un positionnement des attracteurs beaucoup plus systématique pour la catégorie A, qui a des attributs invariants, que pour la catégorie Non-A, où il n'y a pas d'attributs invariants.

Une façon d'expliquer le comportement du réseau peut être avancée en revenant au fait que le FEBAM-SOM a été initialement conçu pour apprendre des stimuli et pouvoir les reconstruire à partir d'input dégradés ou bruités. Le réseau cherche donc à compresser l'information contenue dans les inputs initiaux sur la carte topologique. Nous comptons sur cette compression pour que le réseau ne garde que l'information essentielle concernant les stimuli et qu'ainsi les attracteurs représentent les invariants des catégories formées. Par contre, dans une optique où le but du réseau en est un de reconstruction, il est en effet plus efficace d'accorder une plus grande importance à l'information qui définit la variabilité des stimuli initiaux. Par ailleurs, la question reste entière à savoir pourquoi le réseau a priorisé l'attribut 4, moins variable que l'attribut 5, mais plus variable que les attributs 3 et 1 et 2, en ordre de variabilité respective. Le réseau a été conçu initialement pour traiter des stimuli bidimensionnels. Par contre, rien dans l'architecture du réseau ne laissait présager une problématique liée au traitement de stimuli tridimensionnels. En effet, le seul changement réside dans la dimensionnalité de

la matrice des poids ainsi que dans celle de la carte topologique. Le problème réside donc possiblement dans la structure des stimuli devant être traités par le réseau. Comme expliqué dans la méthodologie, les stimuli ont été générés par la superposition des cartes topologiques des attributs composant les stimuli. Les travaux de Chartier et al. montrent que le réseau traite également toutes les unités d'un stimulus bidimensionnel. Par contre, si le stimulus tridimensionnel est composé de n couches de stimuli bidimensionnels (les cartes topologiques), il est nécessaire de vérifier si le réseau donne la même importance à toutes les couches. Des vérifications ont été faites à partir d'inputs très simples composés de deux couches de dimension 42×4 (donc des tenseurs de dimension $2 \times 4 \times 4$) avec 4 ou 8 unités activées par couche où une couche présentait des activations invariantes alors que l'autre présentait de la variabilité (Voir Figures 11 et 12 en Annexe 2). Lorsque ces stimuli sont présentés au réseau, il donne priorité à la couche présentant la variabilité pour catégoriser les items, indépendamment de la couche dans laquelle ceux-ci se trouvent. Or, nos simulations de la section précédente indiquent que le réseau utilise plusieurs couches. Entre des inputs de taille $4 \times 4 \times 2$ et ceux de taille $20 \times 20 \times 5$ se trouve un changement qualitatif quant à l'information examinée. Il doit donc y avoir un seuil critique, soit de complexité de l'activation ou de la taille de l'input, à partir duquel le réseau commence à accorder de l'importance à plus d'une couche.

Une dernière vérification visait à observer l'influence du nombre de couches dans l'input afin de vérifier si la structure de l'input est ce qui a empêché le réseau de créer un patron d'attracteurs clairs. Des cinq attributs (et donc cinq couches) des stimuli présentés au réseau pour la tâche de catégorisation par règle conjonctive, trois ont été conservées : les deux attributs invariants formant la règle et l'attribut 4 ayant été principalement utilisé par le réseau lors de sa catégorisation. Comme on peut le voir sur la Figure 19 de l'Annexe 2, le réseau catégorise de façon entièrement basée sur l'attribut 4 et l'effet stabilisant des attributs invariants est encore présent. Somme toute, la catégorisation a été effectuée de la même façon avec trois ou cinq attributs.

Modélisation future

Afin d'obtenir un comportement du réseau correspondant au comportement des sujets humains, nous pensons qu'il sera nécessaire d'ajouter un deuxième niveau au

réseau. Ce second niveau servirait à détecter la présence d'invariants afin de diriger le processus catégoriel afin de séparer catégorisation implicite et explicite et agirait à la façon d'un biais attentionnel. Un processus de biais attentionnel est présent dans plusieurs modèles de catégorisation (mais pas dans le FEBAM). Par exemple, ALCOVE (Kruschke, 1992) et plusieurs autres modèles possèdent ce biais programmé par un réseau connexionniste simple et les dernières versions du GCM (Nosofsky, 2002) possèdent aussi un poids attentionnel modélisé par un paramètre libre.

Nous proposons une façon simple d'implémenter un biais attentionnel à l'aide d'un compteur qui serait sensible à la récurrence d'une valeur dans un attribut. Il s'agirait d'établir un seuil au-dessus duquel le réseau doit considérer la récurrence comme signifiant la présence d'un invariant, par exemple une valeur présente la moitié du temps. Cela impliquerait malheureusement de devoir s'éloigner de l'idée d'un réseau auto apprenant parfait puisque le seuil est un paramètre arbitraire décidé par l'expérimentateur et non pas dirigé par les données (cependant, le réseau de course en parallèle possède une règle d'apprentissage pour les seuils de fréquences; Cousineau, Lacroix & Hélie, 2003). Le deuxième niveau permettrait de pondérer l'importance des attributs, laissant l'apprentissage libre lors de l'absence d'invariant, selon un type de catégorisation implicite, et en forçant la catégorisation selon les invariants lorsque ceux-ci sont détectés, selon un type de catégorisation par règle explicite. Par désir de parcimonie, les règles commenceraient par une règle de présence d'un attribut pour se complexifier si la règle est sous-optimale. Aussi, de la même façon que le nombre de catégories a été fourni aux sujets humains, il faudrait donner au réseau ce nombre de catégorie de façon à lui permettre d'optimiser sa catégorisation. Cette suggestion permettrait possiblement au FEBAM-SOM d'émuler les deux types de traitement du COVIS en n'utilisant qu'une seule architecture. Similairement, il permettrait peut-être au réseau de réaliser de l'inférence par air de famille autant que de l'inférence bayésienne, une distinction considérée comme essentielle pour comprendre le traitement sémantique (modèle CLARION, Hélie & Sun, 2010).

Une façon de pallier au problème des attracteurs semblables mais non identiques dû à la recherche de variabilité du réseau serait d'ajouter une couche supplémentaire au

réseau ayant un nombre d'unités égal au nombre de catégories. Par une connexion de type *many-to-one*, il serait possible de connecter des cartes topologiques semblables mais pas identiques à chaque catégorie et ainsi permettre un regroupement qui compense pour la stratégie de conservation de la variabilité du réseau d'avoir une meilleur adéquation avec les résultats empiriques.

La maîtrise étant trop courte, nous explorerons ces possibilités au cours des prochaines années.

Conclusion

Les buts de cette expérience étaient de a) départager au niveau de la difficulté les tâches explicites de catégorisation par règle logique et les tâches de catégorisation implicites par intégration d'information (air de famille) et de b) modéliser les résultats empiriques obtenus à l'aide d'un réseau neuronal. Il s'agit d'une réussite partielle. Nous avons en effet réussi à départager la tâche de catégorisation par règle logique disjonctive comme la plus difficile tel que mesuré par le temps nécessaire pour apprendre la tâche et la capacité de transférer la règle apprise. Ce résultat était attendu. Ce qui est plus intéressant est la comparaison entre la tâche de catégorisation par règle conjonctive et la tâche de catégorisation par air de famille. La tâche implicite a été mieux apprise et plus rapidement que la tâche explicite par règle conjonctive. Les gens auraient donc plus de facilité à intégrer les valeurs possible de cinq attributs et leur relation que d'apprendre une règle rigide de deux attributs parmi cinq. Il serait intéressant de voir la différence de difficulté entre la tâche d'intégration d'information et une tâche explicite où la règle est la présence d'une valeur sur un attribut. Nous avons aussi obtenu les effets de la typicité relevés dans la documentation. Par contre, nous avons obtenu un résultat intéressant : l'effet de familiarité que nous avons trouvé n'est pas indépendant de la tâche à accomplir. En effet, plus une tâche explicite est facile et moins les gens se servent de la familiarité pour leur catégorisation mais lorsque la règle devient trop difficile à apprendre, les gens se rabattent sur la familiarité pour compléter la tâche de catégorisation. Cette familiarité,

qui ne permet pas une catégorisation parfaite, permet quand même une réponse approximativement correcte dans nos tâches (75% pour la haute familiarité, 50% pour la moyenne familiarité). Les participants utilisent d'abord la familiarité la plus évidente initialement avant d'utiliser la familiarité moins forte.

Il reste encore beaucoup à faire pour avoir un modèle de la catégorisation qui ait une bonne correspondance avec le processus catégoriel humain. Tel quel, le réseau FEBAM-SOM est apte à faire une catégorisation par intégration d'information mais ne peut faire de catégorisation avec une règle explicite à moins de lui ajouter un niveau supplémentaire qui agirait comme un biais attentionnel pour diriger le type de catégorisation et la priorisation des attributs à utiliser pour les tâche de catégorisation avec une règle explicite.

Références

- Anderson, J. R. (1991). The adaptive nature of human categorization, *Psychological Review*, 98, 409-429.
- Ashby, F. G., Alfonso-Reese, L. A., Turken, U. & Waldron, E. M. (1998). A Neuropsychological Theory of Multiple Systems in Category Learning. *Psychological Review*, 105(3), 442-481.
- Ashby, F. G. & Maddox, W. T. (1993). Relations between Prototype, Exemplar, and Decision Bound Models of Categorization, *Journal of Mathematical Psychology*, 37(3), 372-400.
- Ashby, F. G. & Maddox, W. T. (2005). Human Category Learning. *Annual Review of Psychology*, 56, 149-178.
- Ashby, F. G. & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, 93, 154-179.
- Ashcraft, M. H. (1978) Property dominance and typicality effects in property statement verification. *Journal of Verbal Learning and Verbal Behavior*, 17, 155-164
- Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11, 629-654.
- Barsalou, L.W. (1987). The instability of graded structure: Implications for the nature of concepts. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 101-140). Cambridge, Cambridge University Press.
- Barsalou, L. W., Huttenlocher, J. & Lamberts, K. (1998) Basing categorization on individuals and events, *Cognitive Psychology*, 36, 203-272.

- Bornstein, M. H., (1987) Perceptual categories in vision and audition. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 287-300). New York, Cambridge University Press.
- Bornstein, M. H., Kessen, W. & Weiskopf, S. (1976) Color vision and hue categorization in young human infants, *Journal of Experimental Psychology: Human Perception and Performance*, 2, 115-129
- Chartier, S. & Boukadoum, M. (2006). A Sequential Dynamic Heteroassociative Memory for Multistep Pattern Recognition and One-to-Many Association, *IEEE Transactions on Neural Networks*, 17(1), 59-68.
- Chartier, S., Boukadoum, M., & Amiri, M. (2009). BAM Learning of Nonlinearly Separable Tasks by using an Asymmetrical Output Function and Reinforcement Learning, *IEEE Transactions on Neural Networks*, 20(8), 1281-1292 .
- Chartier S., Renaud, P. & Boukadoum, M. (2008). A Nonlinear Dynamic Artificial Neural Network Model of Memory, *New Ideas in Psychology*, 26(2), 252-277.
- Chartier, S., Giguère, G. & Langlois, D. (2009). A new bidirectional heteroassociative memory encompassing correlational, competitive and topological properties. *Neural Networks*, 2009 Special Issue, 568-578.
- Cohen, B. & Murphy, G. L. (1984). Models of Concepts. *Cognitive Science*, 8, 27-58.
- Cousineau, D. (2009, novembre). *Une théorie générale des réseaux connexionnistes*, Communication présentée à la rencontre jointe des laboratoires du CONEC et du VIC, Ottawa.
- Cousineau, D., Lacroix, G. L. & Hélie, S. (2003). Redefining the rules: Providing race models with a connectionist learning rule. *Connection Science*. 15(1), 27-43.
- Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, 127, 107-140.

- Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature*, 407, 630-632.
- Feldman, J. (2003). A catalog of Boolean concepts. *Journal of Mathematical Psychology*, 47, 75-89.
- Giguère, G. (2009). Une structure associative bidirectionnelle d'auto-encodage permettant l'apprentissage et la catégorisation perceptuels (Thèse de doctorat en informatique cognitive ayant obtenu la mention excellent). Université du Québec à Montréal, Montréal.
- Hannah, S. D. & Brooks, L. R. (2009). Featuring familiarity: How a familiar feature instantiation influences categorization. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 63(4), 263-275.
- Hélis, S. & Sun, R. (2010). Incubation, insight, and creative problem solving: A unified theory and a connectionist model. *Psychological Review*, 117(3), 994-1024.
- Homa, D. (1984). On the nature of categories. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 18, pp. 49-94). Oxford, Academic Press.
- Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities, *Proceedings of the National Academy of Sciences*, 79, 2554–2558.
- Kleine, R. E & Kernan, J. B. (1992) Strategies for object identification and confidence: Influence of appearance, amount of context, and familiarity. *Perceptual and Motor Skills*, 75(3, Pt 1), 739-745.
- Kohonen, T. (1984). *Self-organization and associative memory*, New York, Springer.
- Kosko, B. (1988). Bidirectional associative memories, *IEEE Transactions on Systems, Man, & Cybernetics*, 18(1), 49-60.

- Kruschke, J. K. (1992). ALCOVE: An exemplarbased connectionist model of category learning. *Psychological Review*, 99, 22–44.
- Lafond, D., Lacouture, Y. & Mineau, G. (2007). Complexity minimization in rule-based category learning : Revising the catalog of Boolean concepts and evidence for non-minimal rules, *Journal of Mathematical Psychology*, 51, 57-74.
- Lamberts, K. (2000). Information-accumulation theory of speeded categorization. *Psychological Review*, 107(2), 227-260.
- Larochelle, S., Cousineau, D. & Archambault, A. (2006) *The role of definitions in categorization and similarity judgments*. In C. Lefebvre et H. Cohen (Eds.) *Handbook of categorization in cognitive science*, Amsterdam: Elsevier.
- Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, 111, 309-332.
- Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics*, 53, 49-70.
- Maddox, W. T. & Ashby, F. G. (2004). Dissociating explicit and procedural-learning based systems of perceptual category learning. *Behavioural Processes*, 66, 309-332.
- Malt, B. C & Smith, E. E. (1982). The role of familiarity in determining typicality. *Memory & Cognition*. 10(1), 69-75.
- Medin, D. L., & Schaffer, M. M. (1978). A context theory of classification learning. *Psychological Review*, 85, 207–238.
- Medin, D.L. (1989) Concepts and conceptual structure, *American Psychologist*, 44, 1469-1481.

- Minda, J.P. & Smith, J.D. (2001). Prototypes in Category Learning: The Effects of Category Size, Category Structure, and Stimulus Complexity. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 27(3), 775-799.
- Minda, J. P. & Smith, J. D. (2002). Comparing Prototype-Based and Exemplar-Based Accounts of Category Learning and Attentional Allocation. *Journal of Experimental Psychology*, 28(2), 275-292.
- Murphy, G.L. (2004). *The Big Book of Concepts*. Cambridge, MIT Press.
- Neisser, U. & Weene, P. (1962). Hierarchies in concept attainment, *Journal of Experimental Psychology*, 64(6), 640-645.
- Nosofsky, R. M. (1988). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(1), 54-65.
- Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, 104(2), 266-300.
- Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, 101(1), 53-79.
- Nosofsky R. M & Zaki S. R., Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 28, 924-940.
- Pevtzw, R. & Harnad, S. (1997) Warping Similarity Space in Category Learning by Human Subjects: The Role of Task Difficulty. In: Ramscar, M., Hahn, U., Cambouropoulos, E. & Pain, H. (Eds.) *Proceedings of SimCat 1997: Interdisciplinary Workshop on Similarity and Categorization*. (pp. 189-195). Department of Artificial Intelligence, Edinburgh University
- Rosch, R. H. (1975) Cognitive reference points, *Cognitive Psychology*, 7, 532-547.

- Rosch, E., & Mervis, C. B. (1975) Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7, 573-605.
- Shepard, R. N., Hovland, C. I. & Jenkins, H. M. (1961) Learning and memorization of classifications. *Psychological Monographs*. 75(13), 42.
- Smith, E. E. & Medin, D. L. (1981). *Categories and concepts*. Cambridge, Harvard University Press.

Annexe 1

Tableaux

Tableau I

Valeur des attributs des stimuli présentés durant la tâche de catégorisation par air de famille

Catégorie	Phase	Attribut 1	Attribut 2	Attribut 3	Attribut 4	Attribut 5	Typicité	Item
A	Apprentissage	1	1	1	1	0	4	1
		2	1	1	1	1		2
		1	1	1	2	1		3
		1	1	2	1	1		4
		1	0	1	1	1		5
		1	1	0	1	1		6
		1	1	1	0	2	3	7
		0	1	1	1	2		8
		2	2	1	1	1		9
		1	0	1	2	1		10
		1	1	0	1	0		11
		0	1	2	1	1		12
	Transfert (valeur jamais rencontrée)	3	1	1	1	1	4	1
		1	1	3	1	1		2
		1	1	1	1	3		3
		1	3	1	1	1		4
		3	1	1	3	1	3	5
		1	1	3	1	3		6
		3	1	1	1	3		7
		1	3	1	3	1		8
		3	1	3	1	3	2	9
		1	3	3	3	1		10
		1	3	1	3	3		11
		3	1	3	3	1		12
	Transfert (valeurs déjà connues)	1	1	1	1	2	4	25
		1	1	0	1	1	4	26
		1	1	1	0	0	3	27
		2	0	1	2	1	2	28
		1	2	1	1	1	4	29
		1	1	1	0	1	4	30
		0	1	2	1	1	3	31
		2	1	0	1	2	2	32
		1	0	0	1	1	3	33
		1	1	2	2	2	2	34
		2	1	1	1	0	3	35
		1	1	1	0	1	4	36

II

B	Apprentissage	0	0	0	0	1	4	13
		2	0	0	0	0		14
		0	0	0	2	0		15
		0	0	2	0	0		16
		0	1	0	0	0		17
		0	0	1	0	0		18
		0	0	0	1	2	3	19
		1	0	0	0	2		20
		2	2	0	0	0		21
		0	1	0	2	0		22
								23
		0	0	1	0	1		23
		1	0	2	0	0		24
								24
	Transfert (valeur jamais rencontrée)	3	0	0	0	0	4	13
		0	0	3	0	0		14
		0	0	0	0	3		15
		0	3	0	0	0		16
		3	0	0	3	0	3	17
		0	0	3	0	3		18
		3	0	0	0	3		19
		0	3	0	3	0		20
		3	0	3	0	3	2	21
		0	3	3	3	0		22
	Transfert (valeurs déjà connues)	0	3	0	3	3		23
		3	0	3	3	0		24
		0	0	0	0	2	4	37
		0	0	1	0	0	4	38
		0	0	0	1	1	3	39
		2	1	0	2	0	2	40
		0	2	0	0	0	4	41
		0	0	0	1	0	4	42
		1	0	2	0	0	3	43
		2	0	1	0	2	2	44
		0	1	1	0	0	3	45
		0	0	2	2	2	2	46
		2	0	0	0	1	3	47
		0	0	0	1	0	4	48

Note: L'ensemble des attributs déterminent l'appartenance à la catégorie A ou à la catégorie B. La typicité est définie comme le nombre d'attributs en commun avec le prototype de la catégorie.

Tableau II

Valeur des attributs des stimuli présentés durant la tâche de catégorisation par règle logique conjonctive

Catégorie	Phase	Attribut					Familiarité		Item
		1	2	3	4	5	Élevée	Moyenne	
A	Apprentissage	1	1	2	3	1	O	O	1
		1	1	2	3	2	O	O	2
		1	1	2	3	3	O	O	3
		1	1	2	1	1	O	N	4
		1	1	2	1	2	O	N	5
		1	1	2	1	3	O	N	6
		1	1	2	2	1	O	N	7
		1	1	2	2	2	O	N	8
		1	1	2	2	3	O	N	9
		1	1	0	3	1	N	O	10
		1	1	1	3	2	N	O	11
		1	1	3	3	3	N	O	12
	Transfert (valeur jamais rencontrée)	1	1	2	3	0	O	O	1
		1	1	2	1	0	O	N	2
		1	1	2	0	0	O	N	3
		1	1	2	0	1	O	N	4
		1	1	2	0	2	O	N	5
		1	1	2	0	3	O	N	6
		1	1	3	3	1	N	O	7
		1	1	1	3	3	N	O	8
B	Apprentissage	0	1	1	3	2	N	N	13
		0	2	2	1	2	N	N	14
		0	3	3	2	3	N	N	15
		1	0	1	3	3	N	N	16
		1	2	2	2	1	N	N	17
		1	3	3	1	1	N	N	18
		2	1	1	3	2	N	N	19
		2	0	2	2	2	N	N	20
		2	3	3	1	1	N	N	21
		3	1	1	3	1	N	N	22
		3	0	2	2	3	N	N	23
		3	2	3	1	3	N	N	24
	Transfert (valeur jamais rencontrée)	0	0	1	0	2	N	N	9
		0	3	2	1	3	N	N	10
		1	2	3	2	0	N	N	11
		1	3	0	3	1	N	N	12
		2	1	1	0	2	N	N	13
		2	0	2	1	3	N	N	14
		3	2	3	2	0	N	N	15
		3	1	0	3	1	N	N	16

Note: Les attributs 1 et 2 déterminent l'appartenance à la catégorie A

IV

Tableau III

Valeur des attributs des stimuli présentés durant la tâche de catégorisation par règle logique disjonctive

Catégorie	Phase	Attribut					Familiarité		Item
		1	2	3	4	5	Élevée	Moyenne	
A	Apprentissage	1	0	0	2	0	O	O	1
		1	0	0	2	1	O	O	2
		1	0	0	0	2	O	N	3
		1	0	0	1	1	O	N	4
		1	0	1	2	2	N	O	5
		1	0	2	0	0	N	N	6
		0	1	0	2	2	O	O	7
		0	1	0	2	0	O	O	8
		0	1	0	1	0	O	N	9
		0	1	0	2	1	O	O	10
		0	1	0	0	1	O	N	11
		0	1	2	1	2	N	N	12
	Transfert (valeur jamais rencontrée)	1	0	0	0	3	O	N	1
		1	0	0	1	3	O	N	2
		1	0	0	2	3	O	O	3
		1	0	0	3	3	O	N	4
		1	0	0	2	2	O	O	5
		1	0	3	2	3	N	O	6
		0	1	0	0	3	O	N	7
		0	1	0	1	3	O	N	8
		0	1	0	2	3	O	O	9
		0	1	0	3	3	N	N	10
		0	1	3	2	3	N	O	11
		0	1	3	2	1	N	O	12
	Transfert (valeurs déjà connues)	1	0	0	1	0	O	N	25
		1	0	0	1	1	O	N	26
		1	0	0	0	1	O	N	27
		0	1	2	2	1	N	O	28
		0	1	0	0	2	O	N	29
		0	1	1	0	0	N	N	30
		1	0	1	1	1	N	N	31
		1	0	1	2	1	N	O	32
		1	0	0	0	0	O	N	33
		0	1	2	2	2	N	O	34
		0	1	2	0	2	N	N	35
		0	1	1	0	2	N	N	36
B	Apprentissage	1	1	2	1	0	O	O	13
		1	1	2	1	1	O	O	14
		1	1	2	0	2	O	N	15
		1	1	2	2	1	O	N	16
		1	1	0	1	2	N	O	17
		1	1	1	0	0	N	N	18
		0	0	2	1	2	O	O	19
		0	0	2	1	0	O	O	20
		0	0	2	2	1	O	N	21
		0	0	2	1	0	O	O	22
		0	0	2	0	1	O	N	23

	0	0	1	2	2	N	N	24
Transfert (valeur	1	1	2	0	3	O	N	13
jamais	1	1	2	2	3	O	N	14
rencontrée)	1	1	2	1	3	O	O	15
	1	1	2	3	3	O	N	16
	1	1	2	1	2	O	O	17
	1	1	3	1	3	N	O	18
	0	0	2	0	3	O	N	19
	0	0	2	2	3	O	N	20
	0	0	2	1	3	O	O	21
	0	0	2	3	3	O	N	22
	0	0	3	1	3	N	O	23
	0	0	3	1	1	N	O	24
Transfert	1	1	2	2	0	O	N	37
(valeurs déjà	1	1	2	0	1	O	N	38
connues)	1	1	2	0	0	O	N	39
	0	0	0	1	1	N	O	40
	0	0	2	0	2	O	N	41
	0	0	1	0	0	N	N	42
	1	1	1	2	1	N	N	43
	1	1	1	1	1	N	O	44
	1	1	2	0	0	O	N	45
	0	0	0	1	2	N	O	46
	0	0	0	0	2	N	N	47
	0	0	1	0	2	N	N	48

Note: Les attributs 1 et 2 déterminent l'appartenance à la catégorie A ou B

Tableau V

Moyennes durant le dernier bloc de la phase d'apprentissage

		Mesure	Tâche de catégorisation par air de famille	Tâche de catégorisation par règle conjonctive	Tâche de catégorisation par règle disjonctive
Dernier bloc de la phase d'apprentissage		Moyenne temps de réponse	1411 ms	1403 ms	1483 ms
		Moyenne taux de bonnes réponses	94,1 %	94,1 %	93,8 %
Ordre de passation	1-2-3	Moyenne temps de réponse	1437 ms	1313 ms	1480 ms
		Moyenne taux de bonnes réponses	92,9 %	93,4 %	93,5 %
	3-2-1	Moyenne temps de réponse	1380 ms	1514 ms	1488 ms
		Moyenne taux de bonnes réponses	95,6 %	94,9 %	94,1 %
		Moyenne temps de réponse	1393 ms	1538 ms	1392 ms
		Moyenne taux de bonnes réponses	93,6 %	94,5 %	94,4 %
Programme d'appartenance	Sciences naturelles	Moyenne temps de réponse	1428 ms	1282 ms	1563 ms
		Moyenne taux de bonnes réponses	94,6 %	93,7 %	93,2 %
	Sciences humaines	Moyenne temps de réponse			
		Moyenne taux de bonnes réponses			

Annexe 2

Figures

Figure 13

Test sur le pattern d'activations 4×4 avec 4 activations par couche

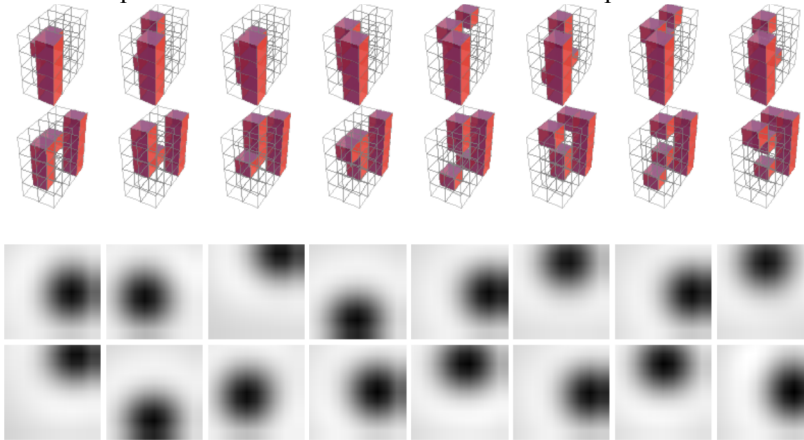


Figure 14

Test sur le pattern d'activations 4×4 avec 8 activations par couche

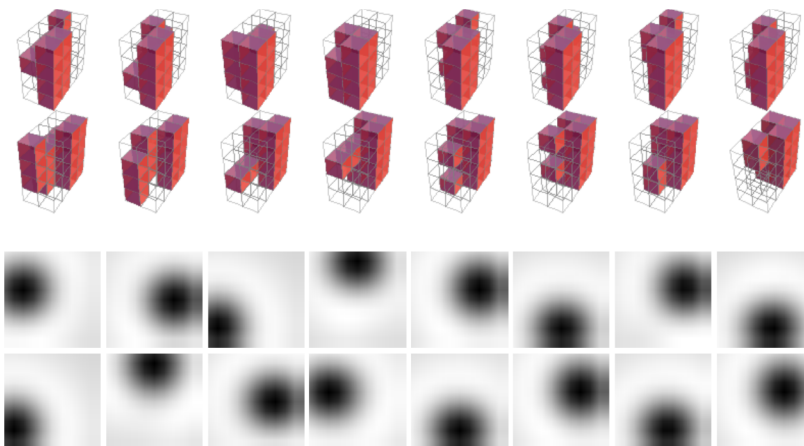
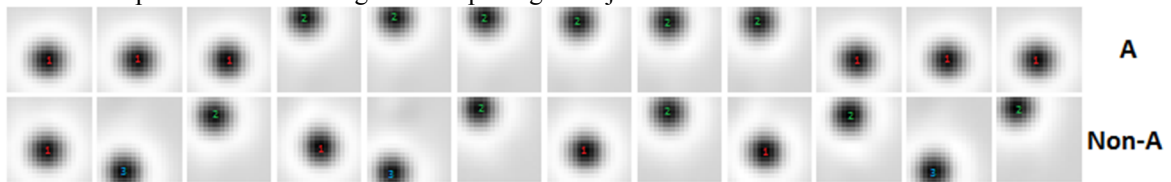


Figure 15

Activations pour la tâche de catégorisation par règle conjonctive avec trois attributs



Annexe 3

Code source pour la création des tenseurs des polyèdres

```

coulP = {0, 1, 2, 3}; polyP = {0, 1, 2, 3}; coulS = {0, 1, 2, 3}; polyS = {0, 1, 2, 3}; dist = {0,
(*Positions: côté, dessus*)
pos = {0, 1};
(*couleurs utilisées: rouge, bleu, vert, jaune*)
convcol = {0 -> {1, 0, 0}, 1 -> {0, 0, 1}, 2 -> {0, 1, 0}, 3 -> {1, 1, 0}};
(*polyèdres utilisés: icosaèdre, octaèdre*)
Forms = PolyhedronData["Platonic"];
convpoly = {0 -> Forms[[4]], 1 -> Forms[[3]], 2 -> Forms[[5]], 3 -> Forms[[1]]};
(*Distances: près, loin*)
convdist = {0 -> {1.5, 0, 0}, 1 -> {2.5, 0, 0}};

Table[
  fig1 = PolyhedronData[polyP[[j]] /. convpoly, "Faces"];
  fig2 = PolyhedronData[polyP[[l]] /. convpoly, "Faces"];
  mat1 = Table[-1, {30}, {30}, {30}];
  print = hi;
  polyhedron = Graphics3D[{
    FaceForm[Opacity[0.7]], Glow[RGBColor[col1 = coulP[[i]] /. convcol]], EdgeForm[Black],
    fig1,
    Opacity[0.7], Glow[RGBColor[col2 = coulP[[k]] /. convcol]], EdgeForm[Black],
    Translate[fig2, posrel = RotateLeft[dist[[m]] /. convdist, pos[[n]]]],
    {Lighting -> None, Boxed -> False}
  ];
  print = hi;
  fig1v = fig1[[2]] /. {x_Integer -> fig1[[1]][[x]]};
  fig1v50 = Round[fig1v[[1]] * 15 + 15];
  poly1 = Flatten[fig1v50, 1];
  tens1 = Table[mat1[[poly1[[i, 1]], poly1[[i, 2]], poly1[[i, 3]]] = +1, {i, 1, Length[poly1]}];
  sz = 5;
  print = hi;
  diff1 = Table[
    If[x == 1 || x == 5 || y == 1 || y == 5 || z == 1 || z == 5,
      1/4,
    If[x == 2 || x == 4 || y == 2 || y == 4 || z == 2 || z == 4,
      1/2,
    1]],
    {x, sz}, {y, sz}, {z, sz}
  ];
  (res = ListCorrelate[diff1, mat1, {sz + 1, 2, 0} // Chop] // MatrixForm;
  Export["vec" <-> ToString[coulP[[i]]] <-> ToString[polyP[[j]]] <-> ToString[coulS[[k]]]
    k <-> ToString[polyS[[l]]] <-> ToString[pos[[n]]] <-> ToString[dist[[m]]] <-> ".mx", mat1],
    {i, 1, 1}, {j, 1, 1}, {k, 1, 1}, {l, 1, 1}, {m, 1, 1}, {n, 1, 1}
  ];
];

```